



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

DEMENTIA AND LOCATION:
LESSONS FROM GEOGRAPHY AND
DATA LINKAGE TO ROUTINELY
COLLECTED DATA

KATHERINE ELIZABETH WALESBY



DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF EDINBURGH

2024



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author. When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

*“Not everything that can be counted counts. . .
and not everything that counts can be counted”*

Albert Einstein

Contents

List of Figures	ix
List of Tables	xi
List of Abbreviations	xiii
Abstract	xviii
Lay summary	xxi
Declaration	xxiv
Funding acknowledgement	xxvii
Acknowledgements	xxviii
Dedication	xliii
Publications, conferences, and thesis related events	xlvi
1 Chapter 1: Introduction	1
1.1 Chapter introduction	1
1.2 Overview of Dementia	1
1.2.1 Dementia definition	1
1.2.2 Dementia prevalence	5
1.2.3 Dementia diagnosis	8
1.3 Overview of ‘big data’	14
1.3.1 Defining ‘big data’ in healthcare	14
1.3.2 Defining ‘routinely collected data’	16
1.4 Overview of dementia epidemiology	17
1.4.1 Life-course epidemiology in relation to dementia	21
1.4.2 Risk factors for dementia	25
1.4.3 Brain health	28
1.4.4 Importance of early/mid-life dementia risk factors	31
1.5 Location, ageing and cognition	35
1.5.1 Defining the terms health and medical geography	35
1.5.2 Location: geography and brain health	36
1.6 Chapter conclusions	51
1.7 Thesis research question, aims and objectives	52
1.8 Overview of PhD	52
2 Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia	55
2.1 Chapter introduction	55

2.2	Research aims and objectives	56
2.3	Published manuscript on <i>medRxiv</i>	56
2.3.1	Systematic review appendices	94
2.4	How does this study contribute to existing literature?	118
2.5	Further discussion since the systematic review	122
2.5.1	Defining an area or ‘place’	122
2.5.2	Location (rural/urban) across our life-course as a risk factor in dementia	124
2.5.3	Location as a risk factor for dementia in different country contexts	128
2.6	Future work	135
2.6.1	Geographic search filters	135
2.6.2	Heterogeneity of studies	137
2.6.3	Risk of bias assessment	139
2.6.4	Publication bias	142
2.6.5	Implications for further research	143
2.7	Chapter conclusions	145
3	Chapter 3: Using big data for dementia research	147
3.1	Chapter introduction	147
3.2	Peer-reviewed published manuscript	148
3.3	Considerations of using routinely collected data for research	156
3.4	Healthcare data: types and coding	160
3.5	Data linkage: unique identifiers, types of linkage, datasets	163
3.5.1	Unique identifiers	164
3.5.2	Types of linkage	168
3.5.3	Datasets: linking to routinely collected data	169
3.5.4	Data sharing	171
3.6	Data governance, access and permissions	173
3.7	Reporting routinely collected data research	178
3.8	Routinely collected data: dementia	180
3.8.1	Death certificates	182
3.8.2	Hospital discharges	183
3.8.3	Prescription of anti-dementia drugs	186
3.9	Routinely collected data: geography	187
3.10	Health data research in relation to ageing and dementia	191
3.10.1	UK health data landscape	191
3.10.2	Current Scottish data strategy	193
3.10.3	Developing ageing health data research UK	197
3.11	Chapter conclusions	200
4	Chapter 4: Investigating dementia in Aotearoa New Zealand (ANZ) using routinely collected data	203
4.1	Chapter introduction	203
4.2	Research aims and objectives	204
4.3	Peer-reviewed published manuscript	205
4.3.1	Supplementary manuscript information – not peer reviewed	214

4.4	Importance of representative data	220
4.5	Why investigate the geographical variation of dementia in ANZ?	226
4.6	Data available in ANZ for epidemiological research	229
4.6.1	Longitudinal datasets and cohorts in ANZ	229
4.6.2	Integrated Data Infrastructure (IDI)	231
4.7	Geographical scales in ANZ	248
4.8	Impact of my research	254
4.8.1	How does this study add to existing literature?	254
4.9	Strengths, limitations and challenges	255
4.10	Additional discussion and future work	255
4.11	Chapter conclusions	259
5	Chapter 5: Methodology for the SMS1947 dementia data linkage study	263
5.1	Chapter introduction	263
5.2	Research aims and objectives	264
5.3	Dementia and the Scottish Mental Survey 1947	265
5.3.1	An overview of Scotland’s population, healthcare and ge- ography	265
5.3.2	Dementia in Scotland	272
5.3.3	The Scottish Mental Survey 1947	273
5.4	Data access and permissions processes	279
5.4.1	Timeline for designing the SMS1947 dementia data linkage study	286
5.5	Study design of the SMS1947 dementia data linkage study	294
5.5.1	Datasets used for dementia ascertainment	294
5.5.2	Creating the study population	302
5.5.3	Merging the flagged datasets	308
5.5.4	Data cleaning and missing data	310
5.5.5	Statistical analysis	311
5.6	Ethics	314
5.7	Chapter conclusions	315
6	Chapter 6: Results of the SMS1947 dementia data linkage study	317
6.1	Chapter introduction	317
6.2	Research aims and objectives	317
6.3	Results	318
6.3.1	Denominator population	318
6.3.2	Numerator population	319
6.3.3	Study population	320
6.3.4	Rural/urban living and dementia	327
6.4	Discussion	331
6.4.1	Comparison with existing literature	332
6.4.2	Reasons for differences between rural and urban living and dementia diagnoses	335
6.4.3	Strengths, limitations and challenges	345
6.5	Future work	358

6.6	Chapter conclusions	362
7	Chapter 7: Discussion	365
7.1	Chapter introduction	365
7.2	Summary of findings	366
7.3	Strengths, limitations and challenges	369
7.4	Impact of my research	376
7.5	Implications of my research	378
7.5.1	What constitutes an urban or rural area?	378
7.5.2	Geography within routinely collected health data	380
7.5.3	Country-specific data for dementia prevalence	384
7.5.4	Accessing routinely collected data	393
7.5.5	Exploring rural vs. urban risk factors for dementia	398
7.6	Future work	401
7.7	Overall conclusions	403
	Appendix 1	407
	Appendix 2	413
	Appendix 3	420
	Appendix 4	421
	Appendix 5	423
	Appendix 6	427
	Bibliography	430

List of Figures

1.1	Explanation of the diagnostic gap via the ‘tip of the iceberg’ analogy	8
1.2	Chronological relationship between pathology, clinical symptoms and biomarkers	13
1.3	The 7 V’s of ‘big data’	15
1.4	Hypothetical model of biomarkers for Alzheimer dementia	20
1.5	Common risk factors in non-communicable diseases	24
1.6	Population level modifiable risk factors contributing to dementia in a life-course model	27
1.7	Stages of brain development across the life-course	29
1.8	Direct and indirect associations: the environment and cognition .	40
1.9	Conceptual model of environmental drivers and unhealthy ageing	41
3.1	Population hierarchy in studies using routinely collected data sources	179
3.2	The privacy-access continuum framework	189
3.3	Scottish Government’s 8 Principles of a Digital Nation	195
4.1	Datasets within the IDI for researchers	232
4.2	The IDI ‘spine’ and different nodes	233
4.3	Outline of the Ngā Tikanga Paihere framework	235
4.4	ANZ map with District Health Boards 2015	252
4.5	Regional variability of health vulnerability across ANZ for physical and mental health conditions	258
5.1	Map of the Scottish Government 8-fold Urban-Rural Classification 2020 system	269
5.2	Data linkage procedure for the SMS1947 Project Two	276
5.3	The HSC-PBPP approval process in Scotland	284
5.4	Flowchart of the process of creating the study population from the original source population	303
5.5	Flowchart depicting the process of dementia flagging in each dataset	307
5.6	Flowchart of process ascertaining dementia from the database population and merging with the study population	310
6.1	Study population: Individuals with and without Dementia by Urban-Rural 2003-2004 Classifications 1-8	325
6.2	Study population: Individuals with and without dementia by SIMD 2004 Deciles	326

List of Tables

1.1	Summary of a selection of key cognitive assessment tools	10
1.2	Conceptual life-course models	22
1.3	Definitions of the four factors in causation	23
1.4	Population attributable fractions (PAFs) and dementia risk	26
1.5	The Six Pillars of Brain Health	30
3.1	Definitions of common governance terms use in routinely collected data	175
3.2	The Five-Safes Framework for research data	176
3.3	The four key domains to improve standardisation in the UK TREs	177
3.4	Population hierarchy in studies using routinely collected data . . .	178
4.1	2016 Longitudinal cohort studies in ANZ	230
4.2	Te Mana Raraunga (Māori Data Sovereignty Network)	238
4.3	Timeline of developing the ANZ IDI dementia project	239
4.4	Electronic datasets used within IDI to ascertain dementia for 2012-2015	240
4.5	Anti-dementia drugs used for dementia ascertainment in ANZ, 2016	242
4.6	ICD 10-AM diagnostic codes used for dementia ascertainment in ANZ	243
4.7	Electronic datasets used within IDI to create annual denominator (<i>study</i>) populations for 2012-2015 (using the IDI October 2016 version)	245
4.8	Geographical scales in ANZ 2015-2017	250
5.1	Geographical scales in Scotland	267
5.2	Deprivation scales in Scotland	270
5.3	The Five-Safes Framework for the <i>SMS dementia data linkage study</i>	280
5.4	Organisations and governance structures in Scotland	282
5.5	Timeline of training undertaken to access data in National Safe Haven	285
5.6	Timeline of planning SMS1947 research in Scotland	293
5.7	Routinely collected health datasets in Scotland for dementia ascertainment in the <i>SMS1947 dementia data linkage study</i>	295
5.8	Anti-dementia drugs used for dementia ascertainment in Scotland	299
5.9	ICD 9 diagnostic codes used dementia ascertainment in Scotland .	300
5.10	ICD 10 diagnostic codes used for dementia ascertainment in Scotland	301

5.11	Population hierarchy in the <i>SMS dementia data linkage study</i> . .	304
5.12	Mock example of grouping, filtering 4 individuals with dementia from records based on first recorded data of dementia within any of the datasets	309
6.1	Datasets where dementia first recorded for the study and database populations	320
6.2	<i>SMS1947 dementia data linkage study</i> population within each Urban-Rural classification	321
6.3	Descriptive statistics of the study population with and without dementia	322
6.4	Cox Proportional Hazard Models (binarized Urban vs. Rural classifications)	328
6.5	Cox Proportional Hazard Models comparing all urban grouped together, against the individual rural classifications	330
A.1	Process for extracting first diagnosis of dementia in SMR	424
A.2	Process for extracting first diagnosis of dementia in NRS mortality	425
A.3	Process for extracting first dementia diagnosis in PIS	426
A.4	Cox Proportional Hazard Models comparing all accessible areas (urban and rural) grouped together, versus all remote areas (urban and rural) classifications.	427

List of Abbreviations

10/66	10/66 protocol for dementia
25OHD	25-hydroxyvitamin D
AAR	Age-Adjusted Rate
ACE	Addenbrooke's Cognitive Examination
ACE-R	Addenbrooke's Cognitive Examination Revised
ADD	Alzheimer's disease dementia
ADLs	Activities of Daily Living
ADR	Administrative Data Research
ADRC	Ageing Data Research Collaborative
ADRD	Alzheimer Disease and Related Disorders
ADRN	Administrative Data Research Network
AI	Artificial Intelligence
amyloid-β	amyloid beta peptide
ANU-ADRI	The Australian National University Alzheimer's Disease Risk Index
ANZ	Aotearoa New Zealand
<i>APOE</i>	Apolipoprotein E gene
<i>APOE-ϵ4</i>	Apolipoprotein E gene epsilon 4 allele
<i>APOE-e4</i>	Apolipoprotein E gene epsilon 4 allele
<i>APP</i>	Amyloid Precursor Protein gene
AUC	Area Under the Curve
BGS	British Geriatrics Society
BMI	Body Mass Index
BNF	British National Formulary
BRNZ	Brain Research New Zealand
CAU	Census Area Units
CBD	Central Business District
CAIDE	Cardiovascular Risk Factors, Ageing and Incidence of Dementia study
CATTs	Consistent Areas Through Time
CCACE	Centre for Cognitive Ageing and Cognitive Epidemiology
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
CFAS	Cognitive Function and Ageing Studies
CHILI	CHI-Indexing and Linkage team
CHI-number	Community Health Index number
CI	Confidence Interval

COSLA	Convention of Scottish Local Authorities
COVID-19	Coronavirus disease 2019
CPH	Cox Proportional Hazards
CT scan	Computerised Tomography scan
CVD	Cardio-Vascular disease
DACHA	Developing research resources And minimum data set for Care Homes Adoption and use study
DAR	Data Access Request
DARE	Data And Research Environments
Da T scan	Dopamine Transporter Scan
DDR	Dementia Diagnosis Rate
DEGURBA	Degree of Urbanisation
DPA	Data Protection Act
DPUK	Dementia Platforms UK
DSM-IV	Diagnostic & Statistical Manual of Mental Disorders Version 4
DSM-V	Diagnostic & Statistical Manual of Mental Disorders Version 5
DNA	Deoxyribonucleic acid
DunedinARB	Dunedin Alzheimer's Disease Related Dementias Risk Benchmark
ED	Emergency department
e-DRIS	Electronic Data Research and Innovation Service
EHR	Electronic Health Records
EIDF	Edinburgh International Data Facility
ENRICH	Enabling Research in Care Homes
EPCC	Edinburgh Parallel Computing Centre
EuGMS	European Geriatric Medicine Society
EWAS	Epigenome-Wide Association Studies
FAIR	Findability, accessibility, interoperability, and reusability principles
FAO	Food and Agriculture Organisation of the United Nations
GBD	Global Burden of Diseases
GDPR	General Data Protection Regulations
GEO	Group on Earth Observations
GHFD	Geolocated Health Facilities Data
GHSL	The Global Human Settlement Layer
GIRFT	Getting It Right First Time
GloDePP	The Global Dementia Prevention Program
GP	General Practice/General Practitioner
GRIPP2	Guidance for Reporting Involvement of Patients and Public version 2
GRADE	The Grading of Recommendations Assessment, Development and Evaluation approach
GRO	General Registrar Office
GUINZ	Growing Up in New Zealand study
GWAS	Genome-wide association studies
HCAP	Harmonized Cognitive Assessment Protocol surveys

HDEC	Health and Disabilities Ethics Committee
HDR UK	Health Data Research United Kingdom
HIC	High Income Country
HPC	High Performance Computing
HR	Hazard Ratio
HSC-PBPP	Health and Social Care Public Benefit and Privacy Panel
IDI	Integrated Data Infrastructure
ILO	International Labour Organisation
IS	Information Scientist
ISD	Information Services Division
IQR	Interquartile range
JRC	Joint Research Centre
LIBRA	Lifestyle for BRAInhealth study
LBC1921	Lothian Birth Cohort 1921
LBC1936	Lothian Birth Cohort 1936
LOAD	Late Onset Alzheimer Disease
LIC	Low Income Country
L-MIC	Low and Middle Income Country
LSOAs	Lower Layer Super Output Areas
LYs	Life-years
MB	Meshblock
MCCD	Medical Certificate of the Cause of Death
MCI	Mild Cognitive Impairment
MDS	Minimum Dataset
MHT	Moray House Test
MMSE	Mini-Mental State Examination
MOOSE	Meta-Analyses of Observational Studies in Epidemiology
MRI	Magnetic Resonance Imaging
MSOAs	Middle Layer Super Output Areas
NC	National Caldicott
NCDs	Non-communicable diseases
NDP	National Digital Platform
NDVI	Normalized Difference Vegetation Index
NICE	National Institute of Clinical Excellence
NIHR	National Institute for Health and Social Care Research
NHI	National Health Index/Indicator
NHS	National Health Service
NHSCR	National Health Service Central Register
NI	Northern Ireland
NIA	National Institute for Ageing
NINCDS-ADRDA	National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (ADRDA)

NINDS-AIREN	National Institute of Neurological Disorders and Stroke (NINDS) and the Association Internationale pour la Recherche et l'Enseignement en Neurosciences (AIREN)
NISRA	Northern Ireland Statistics and Research Agency
NNR	Number Needed to Read
NMDS	National Minimum Dataset
NOS	Newcastle-Ottawa epidemiological observation studies risk of bias
NRS	National Records of Scotland
NSH	National Safe Haven
NT	Northern Territory
NZ	New Zealand
NZ HWR	New Zealand Health, Work and Retirement study
OAs	Output Areas
OECD	Organisation for economic co-operation and development
ONS	Office for National Statistics
OR	Odds Ratio
OPCS-4	Operations and Procedures 4th revision
PAC	Privacy Advisory Committee
PET scan	Positron Emission tomography
PAF	Population Attributable Fraction
PHS	Public Health Scotland
PIS	Prescribing Information System
PM_{2.5}	Particulate matter <2.5 in diameter
PPIE	Patient Public Involvement and Engagement
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRSB	Professional Records Standards Body
PwD	Person with dementia
PwoD	Person without dementia
Python	Python statistical software
R	R statistical software programming
R&D	Research and Development
RDS	Research Data Scotland
REC A	Regional Ethics Committee A
RECORD	REporting of studies Conducted using Observational Routinely-collected data
ROB	Risk of Bias
ROBINS	Risk of Bias Assessment tool for Non-randomized Studies
ROBINS-E	Risk Of Bias In Non-randomized Studies - of Exposure tool
RR	Relative Risk
SA1	Statistical Area 1
SA2	Statistical Area 2
SAIL	Secure Anonymised Information Linkage Databank
SAS	Statistical Analysis System software
SCADR	Scottish Centre for Administrative Data Research

SD	Standard Deviation
SDE	Secure data environment
SES	Socio-Economic Status
SHIP	Scottish Informatics Programme
SIG	Special Interest Group
SILC	Scottish Informatics and Linkage Collaboration
SIMD	Scottish Index of Multiple Deprivation
SMI	Scottish Medical Imaging
SMR	Scottish Morbidity Record dataset
SMR 01	Scottish Morbidity Records 01: General/Acute admission and day case
SMR 04	Scottish Morbidity Records 04: Mental Health inpatient and day case
SMS1932	Scottish Mental Survey 1932
SMS1947	Scottish Mental Survey 1947
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
SNPs	Single Nucleotide Polymorphisms
Snz_uid	Unique identifier for Statistics New Zealand within the IDI
S-PBPP	Statistics Public Benefit and Privacy Panel
SPECT scan	Single-Photon Emission Computerised Tomography scan
STATA	StataCorp statistical software programming
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SQL	Structured Query Language
TBI	Traumatic brain injury
TPDs	Training Programme Directors
TRE	Trusted Research Environment
UAHPEC	University of Auckland Human Ethics Committee
UI	Unique Identifier
UK	United Kingdom
UKSeRP	UK Secure eResearch Platform
UN	United Nations
UNDRIP	United Nations Declaration on the Rights of Indigenous Peoples
UoE	University of Edinburgh
UPI	Unique Patient/Person Identifier
UPRN	Unique Property Reference Number
USRN	Unique Street Reference Number
U-R 8-fold	Urban-Rural 8-fold classification system
VD	Vascular Dementia
VIEW	Vascular Informatics using Epidemiology and the Web
WHO	World Health Organisation

Abstract

Dementia is a growing public health concern, with estimated worldwide prevalence of 152.8 million people by 2050. Dementia prevalence and incidence varies geographically, and this could shed light on risk, and protective, factors for dementia development, or greater understanding if the variation represents differences in accessing a diagnosis for dementia. In this thesis, I use routinely collected data to investigate the association of location - particularly rural and urban living - with dementia.

In [Chapter 1: Introduction](#), I outline my aims and objectives. I give an overview of dementia and dementia epidemiology in relation to my research question: does dementia vary between rural and urban areas and how can we use routinely collected health data to answer this question?

In [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#), I present a systematic review with meta-analysis of observational studies investigating dementia risk with rural versus urban living. Of 13,053 records identified, 545 full-text articles were assessed for eligibility, 38 papers were included in the systematic review, of which 27 could be included in a meta-analysis. Overall, rural living was associated with *increased* odds of dementia, compared to urban living (OR 1.20, 95%CI 1.03-1.40, $p=0.018$). However, when stratifying by country income, this pattern was seen in Low-Middle Income Countries (L-MICs) (OR 1.33, 1.06-1.68, $p=0.014$), but in High Income Countries (HICs) there were *lower* odds of dementia in rural areas compared to urban (OR 1.01, 95% CI 0.88-1.165), $p=0.884$). Future research should consider heterogeneity in the definitions of ‘rural’ versus ‘urban’, as well as growing disparity in methodologies between HIC (large routinely collected data studies) and L-MIC (traditional smaller face-to-face epidemiological studies).

In [Chapter 3: Using big data for dementia research](#), I give an overview of using routinely collected healthcare data (often referred to as ‘big data’) for research, and how these could be used for dementia research. I present my peer-reviewed article on big data in Scotland, and also discuss the health data landscape across the UK. Finally, I consider how future ageing health research using routinely collected healthcare data could expand in the UK.

In the following Chapters ([4](#), [5](#), and [6](#)) I build on the preceding work to present novel research in Aotearoa New Zealand and Scotland using routinely collected data to investigate variation in location and dementia.

In [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#), I investigate the variation of dementia rates between North and South Island in ANZ using a large national data resource, the Integrated Data Infrastructure (IDI), comprising of many routinely collected datasets. The prevalence of dementia in the ≥ 60 year olds varied including between North and South Island (e.g., for 2012, North = 9,696/637,245 people [1.52%], and South = 3,702/230,001 people [1.61%]). Dementia counts were highest amongst the European ethnicity but when age and sex standardised, dementia cases were higher for Māori than non-Māori (2% vs.1%). My methods have since been used by local and international research collaborations investigating dementia in ANZ, and additionally work from other research teams have since linked additional datasets to the IDI to increase dementia ascertainment.

In [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), I outline the processes which were necessary to undertake a research project investigating dementia variation in a historical population level survey, the Scottish Mental Survey 1947 (SMS1947), with individual-level data CHI-linked to routinely collected health data, focusing on investigating rural and urban variation and dementia diagnoses. This chapter highlights the stages of project development, health data linkage methodology, and the challenges encountered. It

outlines the methodology used for [Chapter 6: Results of the SMS1947 dementia data linkage study](#).

In [Chapter 6: Results of the SMS1947 dementia data linkage study](#), I investigate dementia within the Scottish Mental Survey 1947 (SMS1947). Living in a rural region was associated with lower hazard of dementia: HR adjusted for sex, Intelligence score, and Scottish Index of Multiple Deprivation (SIMD) 0.84, 95% CI 0.77-0.93 compared to living in an urban area. Dementia was even lower if living within the most rural/remote region of Scotland (adjusted HR 0.77, 95% CI 0.61-0.98). A major strength of this study was the ability to link a historical national survey (which has a childhood intelligence variable, allowing adjustment for this potential confounder) to healthcare datasets, thus creating a population-level cohort with individual level data from 1947.

In the final [Chapter 7: Discussion](#) I summarise my findings and discuss the impact and implications of my research alongside the strength and limitations. I also outline, and discuss, future research needed and then conclude with final remarks with a focus on the changing landscape of using routinely collected health data as a secondary use in medical research.

Lay summary

Dementia is a growing public health concern, with estimated worldwide prevalence of 152.8 million people by 2050. Dementia prevalence and incidence varies geographically, and this could shed light on risk, and protective, factors for dementia development, or greater understanding if the variation represents differences in accessing a diagnosis for dementia. In this thesis, I use routinely collected data to investigate the association of location - particularly rural and urban living – with dementia.

The opening chapter gives an overview of dementia and highlights the importance of understanding who is at risk of developing dementia and this can help inform things we can do to reduce our future risk of dementia. I introduce my research question for my thesis which is investigating whether dementia varies between rural and urban areas and how we can use data to answer this question.

In the first study ([Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#)), I explore all past research published which has information on developing dementia relating to whether an individual is living in a rural or urban environment. I combine the results to get an overall view on whether living in a rural or an urban environment increases an individual's risk of dementia. I also examine whether that difference varies between High Income Countries (HICs) or Low-Medium Income Countries (L-MICs). My study found that overall rural living was linked with increased chance of dementia compared to urban living.

When undertaking further analysis, I found that there was a difference between HICs and L-MICs which is important to consider. I found that living in a HIC there was a lower risk of dementia in rural areas but living in a L-MIC there was a higher risk of dementia in rural areas. My study builds on past research

investigating geographical location and risk of dementia. I also highlight that future research needs to address changing definitions of ‘rural’ and ‘urban’ within and between countries so that further work can understand the role of location and risk of dementia.

In [Chapter 3: Using big data for dementia research](#), I give an overview of big data in healthcare research, including exploring what data are collected routinely in healthcare and administrative settings and how this can be used by researchers to answer research questions in relation to dementia. I present a narrative overview I have published explaining how data are collected and used in Scotland. Following this, I also discuss the changing landscape of healthcare data research in the UK with a particular focus on how researchers can positively use routinely collected data for ageing health research to improve people’s lives.

The second research study ([Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#)) in my thesis focused on using the nationally collected routine data in Aotearoa New Zealand (ANZ) to investigate how many people had dementia in North and South Island between 2012-2015. It was the first time this routinely collected data source had been used for trying to establish estimates of those with dementia during this time period in ANZ. Estimates of the number of people living with dementia in ANZ previously used information from other countries.

As my study highlights, it is important to use country specific information for estimating the numbers of people with dementia. Countries differ for many reasons and an important conclusion from my study was to be able to create health policies to help those with dementia it is so important to use information relevant to that individual country. My study has since been used by local and international research groups to further understand the numbers of people with dementia in ANZ.

For my third research study in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#), I examined who had dementia and who did not within a historical study from 1947 called the Scottish Mental Survey 1947 (SMS1947). This historical study was linked to routinely collected data and I investigated who had a diagnosis of dementia and where they lived (specifically looking at rural or urban locations). My results showed that those who lived in a rural environment were less likely to have a diagnosis of dementia. This was not influenced by deprivation or childhood cognition (two things that can influence dementia risk). My study builds on my results from [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#) highlighting that living in a rural environment in a HIC is less likely to be linked with a dementia diagnosis.

Interestingly, my results differ from an older historical Scottish Mental Survey from 1932. What my results highlight is that location is important to further explore to understand why this difference exists. It might provide risk factors in our lives that we can modify to reduce our future risk of dementia. I discuss future work that can further explore my findings.

In the final section ([Chapter 7: Discussion](#)) of my thesis, I discuss my overall findings from my research studies and explore how using routinely collected data can be used to further investigate the role of location on our risk of dementia.

Declaration

I declare that the work in this thesis presented for the degree of PhD is an original report of my research, written by me. I confirm that the work is my own, except where work as part of jointly-authored publications has been included. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia

The work presented in Chapter 2 is currently published on *medRxiv* (not peer-reviewed) [1] as ‘Geographical variation in dementia: systematic review with meta-analysis’ by **Katherine E Walesby** (PhD candidate), Susan D Shenkin (PhD supervisor), Jennifer K Burton, Catriona Dunlop, Candida Fenton, Duncan Gray, Luke A McGillicuddy, John M Starr, Tim Wilkinson, Graciella Muniz Terrera and Tom C Russ (principal PhD supervisor). It is being updated with a new search, and including the revised geographic search filters. Following this, it will be submitted to a peer-reviewed paper for publication.

- Conception and design of the study: **KEW**, TCR, JMS and SDS.
- Search strategy design: CF and **KEW**.
- Title and article reviewing: first reviewer **KEW**, second reviewers JKB, TW and reviewer of any conflicts TCR.
- Data extraction: **KEW** with second extraction to cross-check by DG, CD, LAM, TCR and SDS.
- Statistical analysis: **KEW**, SDS, GMT, and TCR.
- Drafted the paper: **KEW**.
- Revised the paper: **KEW**, SDS, JKB, TW, LAM, GMT and TCR.
- Approved the final version of the paper: **KEW**, SDS, JKB, CD, CF, DG, LAM, TW, GMT and TCR.

- Guarantor: **KEW**.
- No others meeting author criteria have been omitted.

I outline the detail of the areas I led on and detailed acknowledgements of the contributions of others in the Acknowledgements Section.

Chapter 3: Using big data for dementia research

The publication included as part of the work presented in Chapter 3 was previously published in *The Journal of the Royal College of Physicians Edinburgh* as ‘What big data could achieve in Scotland’ by **Katherine E Walesby** (PhD candidate), Jennifer K Harrison, Tom C Russ (principal PhD supervisor) [2].

- Conception of the review: **KEW**.
- Design of the review: **KEW**, JKH, TCR.
- Drafted the paper: **KEW**.
- Revised the paper: **KEW**, JKH and TCR.
- Approved the final version of the paper: **KEW**, JKH and TCR.
- Guarantor: **KEW**.
- No others meeting author criteria have been omitted.

I outline the detail of the areas I led on and detailed acknowledgements of the contributions of others in the Acknowledgements Section.

Chapter 4: Investigating dementia in Aotearoa New Zealand (ANZ) using routinely collected data

The work presented in Chapter 4 was previously published in the *Australian Journal of Ageing* as ‘Prevalence and geographical variation of dementia in New Zealand from 2012 to 2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure’ by **Katherine E Walesby** (PhD candidate), Daniel J Exeter, Sheree Gibb, Philip C Wood, John Michael Starr (late principal PhD supervisor), Tom C Russ (current principal PhD supervisor) [3].

- Conception and design of the study: **KEW**, TCR, and JMS.
- Ethics and approvals: **KEW** and DJE
- Creating study population and linking data: **KEW** and SG
- Data extraction: **KEW** and SG.
- Statistical analysis: **KEW** and SG.
- Reviewing data for extraction from safe haven: SG
- Drafted the paper: **KEW**.
- Revised the paper: **KEW**, DJE, SG, PCW, JMS and TCR.
- Approved the final version of the paper: **KEW**, DJE, SG, PCW and TCR.
- Guarantor: **KEW**.
- No others meeting author criteria have been omitted.

I outline the detail of the areas I led on and detailed acknowledgements of the contributions of others in the Acknowledgements Section following.

During my PhD studies, I contributed to a number of other publications, which are referenced throughout the thesis and highlighted in Section ‘Publications, conferences, and thesis related events.’

I declare that this thesis has not been submitted, in whole or in part, in any previous application for a degree.

Katherine Elizabeth Walesby

18th November 2024

Funding acknowledgement

This PhD was funded jointly by Alzheimer Scotland and the University of Edinburgh Centre for Cognitive Ageing and Cognitive epidemiology (CCACE), part of the cross council lifelong Health and Wellbeing initiative (MR/ L501530/1) in the early years. Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. In the subsequent years, funding for continuing my PhD is gratefully acknowledged from Alzheimer Scotland Dementia Research Centre (ASDRC) at the University of Edinburgh.

Additionally, funding from the MRC PhD studentship travel grant in 2016 is acknowledged with thanks for a preliminary visit to New Zealand for planning the research and an additional grant from Alzheimer Scotland to undertake the research in 2017.

The funders played no part in the design or conduct of the research or papers included in this PhD.

Acknowledgements

I would like to firstly acknowledge the contribution a range of people have made to my work in a broader capacity, my clinical-academic mentorship, as well as my well-being over this PhD. Following this, I outline the contribution others have made to specific chapters, pieces of research work or the thesis in general.

This PhD has been a long journey, with a lot of learning personally and professionally. In this journey there have been so many individuals who have nurtured my professional development, supported me during the hurdles, encouraged me and kept me believing it was possible, despite everything in the roller-coaster of life outside of my PhD.

It seems prudent to mention a key quote in data science that I learnt early in my PhD journey about merging an individual's records into a 'Book of Life.' This 'Book of Life' is not in isolation. As individuals we are rarely in isolation, but part of the tapestry of life and connections with others: part of a complex inter-web of interactions with others across that life journey. Halbert Dunn (Chief, US National Office of Vital Statistics) in 1946 described linking an individuals' records:

"Each person in the world creates a Book of Life. This Book starts with birth and ends with death. Its pages are made up of the records of the principal events in life. Record linkage is the name given to the process of assembling the pages of this Book, into a volume."

My 'Book of Life', including this thesis, needs to reflect those that have helped me on this journey. A three-year full-time PhD turned into 2 years full time, a year of maternity leave, and my third year of PhD at less than full time with three extensions and another maternity leave. Nine years in total. A long journey.

And a lot of individuals (supporters, advisors and mentors) along the way. This is to thank them and recognise their kindness and guidance. If I have missed off anyone by mistake, please except my apologies and tiredness!

Before listing a number of people in detail who have guided me, I would like to firstly highlight my family who have had faith that I could finish this and have encouraged me, supported me and been a steady ‘rock’ over the last 9 years to reach this final stage. To my husband Abhishek who said ‘go for it’ at the start and reminded me I can do it when it looked impossible; to my parents who have helped in caring for Anya and Neil as wonderful grandparents as I worked many hours to try and finish the thesis, and for instilling education and learning in me from a young age with determination to overcome hurdles that life can throw at you.

Detailed professional thanks:

To Professor Susan Shenkin and the late Professor John Starr (RIP) for renewing a drive in me to do research when I worked with them clinically in 2009 and later connecting me with Dr Tom Russ who became my principal PhD supervisor. I am grateful to all my supervisors who continue to be exemplary examples of supporting women in clinical academia in medicine, which even in the 21st Century is sadly not always the case. Diversity and inclusivity are words I am even more aware of during my PhD. Your approach to inclusive research and supervision is outstanding. Without it, I would have found it incredibly hard to return after Anya’s diagnosis with so much uncertainty in life. You have the unique qualities of kindness, wisdom and incredible knowledge. You combine it all with humility and dedication to help nurture the next generation of researchers.

Aotearoa New Zealand (ANZ)

My time in ANZ was enriching personally, and a great learning opportunity professionally. Thank you to everyone who made it possible, welcoming me and allowing me to share ideas and learn from your knowledge of dementia and research in ANZ.

Understanding different cultures, and different approaches to life and research is inspiring. The Te reo Māori (language of Māori) has made me reflect as a clinician and researcher on society's approach to disease and disability. The word disabled is *whaikaha* which means to have strength, ability, otherly abled or enabled

Compared with the English language which implies different is less, this culture uses phrases which can empower differences amongst us. I am grateful to the MRC PhD Studentship grant for affording me the opportunity to travel and learn in a different country. Alongside the research I have produced, it has given me much food for thought and life-long reflection.

I am grateful for the initial opportunity to visit Auckland University of Technology and meet with Professor Valery Feigin and Dr Rita Krishnamurthi. Thank you to Dr Sheree Gibb, Dr Phil Wood, and Statistics New Zealand support staff for their assistance, support and insightful conversations sharing their knowledge and support undertaking my research. Thanks to Professor Barry Milne and the COMPASS team at the University of Auckland for their hospitality and allowing me to use their IDI datalab.

Additional thanks to Dr Sarah Cullum for helpful discussions around the differences between dementia care in the UK and ANZ. Meeting with Professor Ngaire Kerse in Auckland was a wonderful opportunity for insightful and inspiring learning from her extensive work investigating dementia in ANZ, along with her warm welcome. Thanks to the many individuals involved in research across ANZ (Pro-

fessor Fiona Alpass, Dr Joanne Allen, Dr Corina Grey, Professor Lynette Tippet, Professor Ian Kirk, Dr Nick Garrett), or at the Ministry of Health (Ross Judge) who gave up their time to have helpful discussions with me about dementia and older people in ANZ. Their time and warm welcome to ANZ is greatly appreciated.

A very special thanks to Professor Daniel Exeter and Ms Heather Seal at the University of Auckland. They welcomed and supported me at the university, and through many life, and work, hurdles in ANZ and Scotland. It is hard to express in words how much their support has meant to me over these many years. I have learnt so much about data science, geography and undertaking research in Aotearoa New Zealand from Professor Exeter. Thank you to both of you for reviewing, and proof-reading, several of the chapters of this thesis. Both of you are unique individuals - incredibly talented people with deeply caring and compassionate personalities. I'm honoured to call them both lifelong friends.

Scotland

Thanks to Professor Ian Deary, Dr Matthew Iverson and the research coordinators at eDRIS for their guidance and support over the years, particularly with the work on the SMS 1947 project. Thanks to Professor Alasdair MacLulich for discussing applying for a clinical research fellowship in geriatric medicine many years ago.

Dr Jennifer Burton - We started our PhD journey together. From colleagues and anxious first year PhD students we became good friends. Your incredible friendship has helped me keep sane through some of the most trying of times personally during this PhD. Your guidance professionally has kept me focused on the end goal of the thesis. Your kindness and compassion helped me in difficult times, and along with Dr Joel Burton made those early year trips to Edinburgh Children's hospital easier with the notes of encouragement and delicious home-made biscuits! Thank you. Without a doubt you are a rising star in academic geriatric medicine research and it has been an honour to work alongside, and

learn from, you in the early PhD years.

Dr Johanna Brown - Jo is not only a professional colleague but a great friend, mentor and confidante. She has been a tower of strength over the recent years during difficult periods. She has also generously given up her time to proof-read the thesis, and always encouraging me at difficult times to believe I can finish this. Thank you.

Finally, I would like to thank a number of people who have helped in different ways over the years and across my career.

In Dundee, my respect and admiration for Dr Carolyn Leslie and Dr Lyn Ramage (RIP), who encouraged me when I was applying for many PhD fellowships in the earlier years, and from whom I learned a lot - both from their clinical acumen and compassion to their patients.

From Dundee to Newcastle, Professor Miles Witham has been a wonderful mentor, supporting me on my research journey. Without a doubt, his encouragement and guidance helped me to apply for this PhD fellowship. I learnt much from his clinical acumen, and using clinical experience to drive forwards research. Later, his compassion helped me again during very difficult times with Abhishek and Anya's health. I won't forget his impromptu and compassionate phone call after he heard of Anya's diagnosis. His measured and wise advice of taking a day at a time was influential in me not giving up on academia following that devastating diagnosis.

Thank you to the many support staff in Alzheimer's Scotland Dementia Research Centre, in particular Denise Munro and Stephanie West. I am honoured to consider you also as friends. Thanks to my past Training Programme Directors (TPDs) Dr Donna Clark and Dr John Harper, for helping me navigate the training aspects in the early years when out of programme for research, and to my

current TPDs Dr Gordon Duncan and Dr Sarah Marrinan for helping navigate a return to clinical training.

To the University of Edinburgh staff, in particular Dr Gordon Duncan as my thesis chair and Professor Paddy Hadoke, Director of Postgraduate Research and Early Career Research Experience. Both have offered compassion and encouragement, alongside some very helpful guidance, during a particularly difficult time with Anya's health near the end of my PhD.

To Dr Richard Dodds (RIP) who kindly helped me prepare for my PhD fellowship interview and was always keen to hear about my research plans from our time on the British Geriatrics Society (BGS) trainees' committee onwards. To Dr Tim Wilkinson, for his time and many helpful discussions about using routinely collected data for dementia research.

Lastly, support comes in many forms in helping us juggle life. Therefore, I am grateful to the Children's Hospice Across Scotland (CHAS) charity for their home nursing support team and Rachel House Children's hospice for creating an environment to celebrate life as well as support and advocate during difficult times. Their nurses and support workers, especially Amanda, Karyn and the two Julies have supported us for the last 6 years and become friends. A special thanks to an amazing teacher Helen McCormack who amongst juggling all her work has given up her Friday afternoons for years as a CHAS volunteer bringing us Friday dinner always with a much needed supportive chat and a listening ear. These little acts of kindness all matter in this journey. Likewise, a thanks to those caring healthcare staff members at Edinburgh Children's Hospital (so many to name individually) who have supported Anya and our family and chosen to deliver true person-centred care. Amongst emergencies, diagnostic uncertainty, and with limited evidence based medicine to guide clinical practice, they have shown a desire to listen to those advocating for a patient's needs whilst recognising that as clinicians we all only know what we know of living with a disease through

learning, professional experience and the interaction we have with patients. Patients, and those family members caring for them, are the experts in living with their disease. They live with a condition 24/7, 365 days a year. This is something I am even more aware of now as a clinician and researcher myself for my future clinical academic career. Amongst a myriad of care, I am grateful to have seen true examples of person-delivered care which I hope to emulate in my future clinical career.

Detailed acknowledgements for each thesis chapter:

Chapter 1: Introduction

I designed and wrote this chapter. My supervisors (Dr Tom Russ and Professor Susan Shenkin) and Professor Daniel Exeter reviewed and commented on it. Their helpful suggestions have been incorporated into the revisions. I am grateful to Ms Heather Seal and Dr Johanna Brown for proof reading it.

Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia

I led on these areas of this research:

- Designing the research question, study design and search strategy
- Drafting the protocol and submitting/amending on Prospero
- Importing references from the search strategy to the online platform Covidence for reviewers to access
- Screening all titles/abstracts and all full text articles and comparing with the second independent reviewer
- Resolving any conflicts in studies for inclusion/exclusion
- Extracting data from all studies (descriptive and quantitative)
- Calculating odds ratios from raw data
- Contacting authors for additional information
- Scoring risk of bias for all studies
- Performing the meta-analysis
- Creating the figures and tables
- Drafting the manuscript, with revisions from co-authors
- Finalising the manuscript and submitting to the journal
- I wrote this chapter incorporating revisions from my supervisors

Acknowledgement of others' contribution to this chapter:

- Dr Tom Russ, Professor Susan Shenkin and the late Professor John Starr

were involved in designing the study, editing the protocol.

- Ms Candida Fenton helped with designing and undertook the search strategy
- Dr Tom Russ and Professor Susan Shenkin also edited the extraction form, the ROBANS risk of bias assessment, that I created and reviewed the extracted data. They also advised on the analysis.
- Dr Jennifer Burton and Dr Tim Wilkinson assisted in reviewing all abstracts and full text articles as second reviewer.
- Dr Tom Russ assisted on resolving any questions regarding article selection.
- Dr Duncan Gray, Dr Luke McGillicuddy and Dr Catriona Dunlop extracted all descriptive data to ensure double extraction (alongside my first extraction).
- Dr Luke McGillicuddy, Dr Tom Russ and Professor Susan Shenkin assisted with scoring risk of bias. Dr Catriona Dunlop assisted with double extraction, or calculation, of odds ratios.
- Dr Graciela Terrera-Muniz advised on statistical questions regarding odds ratios extraction and interpretation.
- All authors reviewed/edited the manuscript apart from Professor John Starr who died in 2018.
- My PhD supervisors (Dr Tom Russ and Professor Susan Shenkin) provided comments on this chapter which I included in my revisions.

Chapter 3: Using big data for dementia research

My contribution to this work:

- I wrote this chapter and incorporated suggestions from my supervisors for revisions.
- Additionally, I led on:
- Developing the idea, drafting and developing the peer-reviewed published review included in this chapter ‘What Big Data could achieve in Scotland?’
- Chairing a 2021 British Geriatrics Society workshop on the role of digital data in the healthcare of older people as part of the Ageing Data Research

Collaboration (ADRC) which generated discussion and the subsequent publication of which I am a co-author.

Acknowledgement of others' contribution to this chapter:

- Dr Jennifer Harrison (nee Burton) and Dr Tom Russ co-authored the published article in this chapter 'What Big Data could achieve in Scotland?'[2]
- Dr Daniel Exeter co-wrote the book chapter with me 'Using big data in primary care research' [4]. Additional thanks for his advice discussing the complexity of geographical, routinely collected data, and for reviewing the chapter with suggested revisions which I have incorporated.
- As part of the Ageing Data Research Collaboration (ADRC), I wish to thank the members of the collaboration for working with them as a co-author on several published papers, for their contribution and discussion to the British Geriatrics Society workshop in 2021 on the role of digital data in healthcare of older people, and the statement of thoughts submitted collectively from ADRC to The Sudlow Review. All of these contribute to the themes of this chapter.

Chapter 4: Investigating dementia in Aotearoa New Zealand (ANZ) using routinely collected data I led on these areas of work:

- Identifying existing datasets in ANZ and assessing their suitability for the research question
- Establishing a collaboration with researchers in ANZ
- Meeting with a number of researchers to discuss the IDI and other cohorts available during a visit in 2016
- Initiating a meeting with a New Zealand Ministry of Health representative in Wellington to discuss and understand dementia recording in ANZ during a visit in 2016
- Initiating, contacting and planning meetings and follow up with Statistics NZ to discuss the research question and study design during a visit in 2016
- Designing the research question and study design

- Writing the protocol
- Gaining approvals from Statistics NZ and the University of Auckland ethics committee
- Organising two visits to ANZ to meet researchers and subsequently undertake the analysis within the data lab (in 2016 and 2017), working remotely with a statistician (Dr Sheree Gibb) in Wellington.
- Working with the Dr Gibb to adapt the SAS code in the IDI
- Performing initial data analysis and interpretation
- Working with the Dr Gibb and Statistics NZ to resolve any issues
- Working with the Dr Gibb when I was back in Scotland (therefore unable to access the data within the safe haven) to discuss the analysis that I continued to lead, but was undertaken by her under my instruction
- Creating the figures and tables from the outputted data
- Drafting the manuscript, with revisions from co-authors
- Finalising the manuscript and submitting to the journal
- Writing the poster for presentation at international conferences
- Writing this chapter including revisions from my supervisors

Acknowledgement of others' contribution to this chapter:

- I am grateful for the many discussions about the project and revisions of the manuscript with my co-authors Dr Sheree Gibb, Professor Daniel Exeter, Dr Phil Wood, Dr Tom Russ and the late Professor John Starr.
- Thanks to Dr Jingeng Zhao and Dr Sheree Gibb for allowing me to adapt their SAS code within the IDI that was developed for their VIEW research project.
- Thanks to Professor Daniel Exeter for helping navigate the ethics processes at the University of Auckland, for supervising me at the University of Auckland during the research project in 2017, his introduction to Drs Zhao and Gibb, and for his helpful discussion and guidance on the geographical scales available in ANZ.
- Thanks to Dr Sheree Gibb, statistician, who helped ensure we created the

correct denominator population and assisted with SQL which I was not familiar with. Her advice and help were instrumental in creating the correct code to extract people with dementia from the different datasets. She reviewed and amended the SAS code to ensure accurate extraction of people with dementia and creating an accurate denominator population within the IDI. Lastly, she ensured the data passed statistical disclosure control for outputting from the IDI data lab when I could not be in ANZ working in the data lab.

- Thanks to Dr Phil Wood for his discussion on the potential to use memory clinic data, the prescribing of anti-dementia drugs, and overall understanding of dementia care and practices in ANZ.
- Thanks to Jessie Colbert, University of Auckland, for her time and skills creating the map of Auckland for me in this chapter.

Chapter 5: Methodology for the SMS1947 dementia data linkage study

My contribution to this work:

I designed the research methodology and wrote this chapter. I am grateful to my supervisors for their comments, which I have incorporated in modifying this chapter.

Acknowledgement of others' contribution to this chapter:

- I am also grateful to the many eDRIS research co-ordinators who helped with both the first and final research project. In particular, I would like to thank the current research co-ordinator Hannah Mackenzie.
- I am grateful to the HSC-PBPP panel for their time in reviewing my project and approval, and also the ethics committees.
- I acknowledge Professor John Starr (deceased) and Professor Ian Deary for their role in safeguarding the valuable SMS1947 ledgers, allowing the data linkage to be possible. Their lifetime professional dedication to ensuring the Scottish Mental Surveys were digitally linked enabled me to undertake this research.

- Thanks also to Dr Matthew Iveson for his assistance having worked on the SMS1947 CHI-linked data within Scotland's National Safe Haven. His help was invaluable in navigating how best to undertake the analysis. Thank you for giving a lot of your time discussing some important aspects of the SMS1947 and related methodology and reviewing my adapted R code.
- Thanks to Dr Matthew Iverson and Professor Ian Deary for submitting the minor amendment including a request to extend the project deadline and additional geographical data for my research.
- A final thanks to the many teachers and other professionals involved in developing and organising the age 11 test in 1947 across Scotland and importantly the eleven year olds who undertook that test allowing me the opportunity to undertake my research almost over 70 years later.

Chapter 6: Results of the SMS1947 dementia data linkage study My contribution to this work:

- Designed the research methods for using the SMS1947 CHI linked data within the national safe haven.
- I adjusted R code from previous projects and created new code, and checked this with another researcher within the national safe haven, Dr Matthew Iveson.
- I undertook the ascertainment of dementia within the datasets in Scotland's National Safe Haven (NSH), flagging everyone with a diagnosis of dementia.
- I created the study population from the database population within the NSH.
- I adapted previous R code, and wrote new code, to undertake the Cox proportional hazards statistical models in the NSH and the descriptive results table.
- I extracted the information required for this results chapter and wrote the disclosure forms to have the data checked prior to output from the national safe haven by the team.
- I wrote this chapter, created the tables and figures and incorporated sug-

gestions and revisions from my supervisors (Dr Tom Russ and Professor Susan Shenkin).

Acknowledgement of others' contribution to this chapter:

- I am grateful for the assistance of Dr Matthew Iveson for his knowledge and expertise and helpful discussions to help in my understanding of the SMS1947 CHI-linked health data within the National Safe Haven.
- Additionally, I am grateful for Dr Matthew Iveson's assistance and discussion for developing R code to create the study population and reviewing my updated R code for the analysis.
- To Drs Matthew Iverson, Emily Ball and Drew Altschul for allowing me to adapt and modify their previous R code used for a different project on the SMS1947 CHI-linked health data within the National Safe Haven.
- To my supervisors Dr Tom Russ and Professor Susan Shenkin for discussing the statistical analysis and best methods to undertake it, and for Dr Tom Russ reviewing and assisting with the R code for this.
- To my supervisors for reviewing this chapter and their helpful comments which I have included in the amendments.
- To the eDRIS research coordinator teams for Scotland's National Safe Haven, particularly Hannah Mackenzie for reviewing and submitting my amendments requesting additional dementia deaths within the dataset and for reviewing and authorising the outputs of the research for outside the National Safe Haven.
- To Professors Ian Deary and the late John Starr for their early discussion in developing this project, and their support over the years.

Chapter 7: Discussion

I designed and wrote this chapter. I am grateful to my supervisors Dr Tom Russ and Professor Susan Shenkin for their helpful comments that I included in the revisions of this chapter.

Acknowledgement of additional thanks

To format my thesis I used Latex in overleaf. I am grateful to Ashley Smith (University of Edinburgh, UoE) for publicly sharing an open access thesis template in 2019 for latex using overleaf in UoE thesis structure. I have been able to use, expand and adapt this template for my thesis formatting.

A special thanks to my sponsors over the years (initially Alzheimer Scotland and MRC funded Centre for Cognitive Ageing and Epidemiology (CCACE) and in the later years via Alzheimer Scotland Dementia Research Centre, ASDRC) who have enabled me to undertake this PhD fellowship, and for accommodating my extensions due to maternity leave and my daughter's health. My initial sponsorship via Medical Research Council and ASDRC for my initial visit to Aotearoa New Zealand. Thanks to the University of Edinburgh and Universities in Aotearoa New Zealand (Auckland University of Technology and University of Auckland) and additional thanks for the grants via Medical Research Council PhD studentship travel grant and Alzheimer Scotland to undertake my research in Aotearoa New Zealand.

Finally, I would like to thank the 11 year olds who undertook the Moray House Test in 1947 as part of the Scottish Mental Survey, and the many professionals who coordinated this, allowing me to research dementia in this anonymous linked dataset over 7 decades later. Additionally, I thank the many individuals (public and patients) who support the safe development of health data research to investigate diseases such as dementia, without which the field of health data science could not progress and research advancements using this valuable resource would not be possible.

Dedication

At the start of my PhD in 2015, my supervisor (the late Professor John Starr) gave some advice. That advice I have reflected on much during the years of my PhD. He said, a PhD is not only about producing research and publishing. A PhD is about the journey, the training, and the skills to develop as a clinical academic and about enabling you to become an independent researcher post PhD. His sudden death in 2018 was a great loss and he is greatly missed.

That ‘journey’ has been long for me. Professionally, as with every PhD, there are hurdles to navigate and it’s been a journey of self-development and academic learning during a time of change in the area I have been studying. However, some of the greatest self-learning has come from personal challenges which I could not have even remotely envisaged starting out in 2015. At times, I was unclear how I could continue to navigate balancing academic work and motherhood as a carer to a child with an ultra-rare complex life-limiting disease.

Therefore, I dedicate this PhD to my children, Anya and Neil. In your different ways you have both inspired me, and reminded me how fragile life can be. Amongst all the sadness and challenges, you remind me of the innocence of children with a love of life, a twinkle in your eyes, and laughter that lights up my world, which keeps me going in darker times.

Anya darling, my hope is that there are researchers writing PhDs in the future with breakthroughs in your cruel disease. You endure so much pain and uncertainty. You inspire me for your ‘zest for life’ despite everything you have been through in your short few years. You remind me that you don’t need to speak to be able to communicate, and showed me how society needs to change to truly be inclusive to all with disabilities and additional needs. There are times during this PhD journey that we were close to losing you, but your determination for life has been motivation for me to keep going. You are a beacon of hope and

light, a determined little girl who does not let the obstacles of paralysis, pain, seizures, life-threatening events and disability stop you from reaching your goals whether that's trying to crawl, walk, or say 'mama.' For this your inner strength inspires me. I wrote one book chapter in Edinburgh Sick Kids' hospital watching you while you slept and recovered after an emergency. It will always be a special book chapter in my heart, written during a painful time. Through many obstacles I have prepared this thesis, and when the road ahead seems hard you and your brother's smile have driven me to persevere. This is for you and your amazing caring, and compassionate younger brother.

Finally, in recognition of the learning and life-long friendship I made during my research in Aotearoa New Zealand, I include a Māori proverbial saying [5]:

Whāia te iti kahurangi, ki te tuohu koe, me he maunga teitei

It translates as: *"Seek the treasure you value most dearly: if you bow your head, let it be to a lofty mountain"* [5]. It is about perseverance, endurance and overcoming hurdles, something inspiring to me in this long and challenging PhD journey.

Publications, conferences, and thesis related events

Peer-reviewed publications presented in this thesis:

- **Walesby KE**, Exeter DJ, Gibb S, Wood PC, Starr JM, Russ TC. Prevalence and geographical variation of dementia in New Zealand from 2012 to 2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure. *Australas J Ageing*. 2020 May 11. doi: 10.1111/ajag.12790 [3].
- **Walesby KE**, Harrison JK, Russ TC. What big data could achieve in Scotland. *J R Coll Physicians Edinb*. 2017 Jun;47(2):114-119. doi: 10.4997/JR-CPE.2017.201 [2].

Non peer-reviewed publications presented in this thesis:

- **Walesby KE**, Shenkin SD, Burton JK, Dunlop C, Fenton C, Gray D, McGillicuddy LA, Starr JM, Wilkinson T., Muniz Terrera G., Russ, TC. Geographical variation in dementia: systematic review with meta-analysis. medRxiv 2023.12.11.23299178; doi: 10.1101/2023.12.11.23299178 [1].

Publications referred to in the thesis (in appendices):

- Masoli JAH, Todd O, Burton JK, Wolff C, **Walesby KE**, Hewitt J, Conroy S, van Oppen J, Wilkinson C, Evans R, Anand A, Hollinghurst J, Bhanu C, Keevil VL, Vardy ERLC. New horizons in the role of digital data in the healthcare of older people. *Age Ageing*. 2023 Aug 1;52(8):1-7 [6].
- Todd OM, Burton JK, Dodds RM, Hollinghurst J, Lyons RA, Quinn TJ, Schneider A, **Walesby KE**, Wilkinson C, Conroy S, Gale CP, Hall M, Walters K, Clegg AP. New Horizons in the use of routine data for ageing research. *Age Ageing*. 2020 Aug 24;49(5):716-722. doi: 10.1093/ageing/afaa018 [7].
- Doney ASF, Bonney W, Jefferson E, **Walesby KE**, Bittern R, Trucco E, Connelly P, McCrimmon RJ, Palmer CNA. Investigating the Relationship

Between Type 2 Diabetes and Dementia Using Electronic Medical Records in the GoDARTS Bioresource. *Diabetes Care*. 2019 Oct;42(10):1973-1980. doi: 10.2337/dc19-0380 [8].

- Exeter D, **Walesby KE**. Using big data in primary care research; Oct 2018, Book Chapter in *How to do Primary Care Research*. CRC press [4].
- Starr JM., **Walesby KE**. Diagnosis and Management of Dementia in older people. *Medicine* 2017; 45(1): 51-54 [9].
- Starr JM., **Walesby KE**. Vascular and mixed dementia. Book Chapter in *Encyclopedia of Gero-psychology*. Editor / Nancy. A. Pachana. Springer, 2016 [10].

Other papers published during the course of this thesis:

- Welch C, Wilson D, Sayer AA, Witham MD, Jackson TA. Development of a UK core dataset for geriatric medicine research: a position statement and results from a Delphi consensus process. *BMC Geriatr*. 2023 Mar 23;23(1):168. doi: 10.1186/s12877-023-03805-5 [11] (collaborator as part of the UK Geriatric Medicine Core Dataset Extended Working Group).
- Lim S, Cox NJ, Hale M, White L, Welch C, Lochlainn MN, Burton JK, Richardson SJ, **Walesby KE**, Willott R, Makin S, Taylor J. Engagement in research during specialist geriatric medicine training: results of a national trainee survey. *Clinical Medicine*. 2022 Nov 22(6):553-558 [12].
- Caulfield L, Heslop P, **Walesby KE**, Sumukadas D, Sayer AA, Witham MD. Effect of Angiotensin System Inhibitors on Physical Performance in Older People – A Systematic Review and Meta-Analysis. *J Am Med Dir Assoc*. 2020 Aug 25:S1525-8610(20)30623-X. doi: 10.1016/j.jamda.2020.07.012 [13].
- Welch C, Geriatric Medicine Research Collaborative. Growing research in geriatric medicine: a trainee perspective. *Age Ageing*. 2020 Aug 24;49(5):733-737. doi: 10.1093/ageing/afaa052 [14] (collaborator author as part of the Geriatric Medicine Research Collaborative).
- Burton JK, Ferguson EEC, Barugh AJ, **Walesby KE**, MacLulich AMJ,

Shenkin SD, Quinn TJ. Predicting Discharge to Institutional Long-Term Care After Stroke: A Systematic Review and Metaanalysis. *J Am Geriatr Soc.* 2018 Jan;66(1):161-169. doi: 10.1111/jgs.15101 [15].

- Harrison JK, **Walesby KE**, Hamilton L, Armstrong C, Starr JM, Reynish EL, MacLulich AMJ, Quinn TJ, Shenkin SD. Predicting discharge to institutional long-term care following acute hospitalisation: a systematic review and meta-analysis. *Age Ageing.* 2017 Jul 1;46(4):547-558. doi: 10.1093/ageing/afx047 [16].
- Simpkin AL, **Walesby KE**. Training tomorrow's doctors. *Future Hosp J.* 2017 Feb;4(1):56-60. doi: 10.7861/futurehosp.4-1-56 [17].
- **Walesby KE**, Lyall MJ, Mackay TW, Wood B, Bell D. Valuing our trainees: the future of medicine in the UK. *J R Coll Physicians Edinb.* 2016 Sep;46(3):146-149. doi: 10.4997/JRCPE.2016.301 [18].
- **Walesby, Katherine.** Updates in General Internal Medicine. *J R Coll Physicians Edinb.* 2016; pp. 1-2.

Key conference and meeting presentations:

- April 2024: 3 Minute Thesis oral presentation, University of Edinburgh CMVM
- July 2022: Poster presentation at Alzheimer Association International Conference (AAIC) *Systematic review of geographical variation of dementia*
- December 2021: Oral presentation at Scottish Neuroprogressive and Dementia Network Conference. *Rural and urban differences in dementia*
- November 2021: Chaired three speakers at British Geriatrics Society Autumn Conference on 'Big Data Adapting Hospital Electronic Health Records (EHR) to improve the care of Older people' and chaired subsequent Debate and discussion (speakers and conference attendees) on Big Data.
- June 2021: Led successful submission (and then chaired three speakers and conference discussion) a conference workshop for Royal College of Psychiatrists International Congress. *Life course psychiatric research: New horizons, new perspectives and new directions.*

- June 2021 (delayed from June 2020 due to COVID-19): Virtual Poster presentation at Nordic Congress on Ageing *Prevalence and geographical variation of dementia in New Zealand (NZ) from 2012-2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure (IDI)*
- March 2021: Invited dementia expert panellist for Royal Lyceum Edinburgh's post-show discussion on dementia and the play 'Angela' by Mark Ravenhill
- July 2020: (postponed due to COVID-19) Poster presentation at Alzheimer Association International Conference (AAIC) *Prevalence and geographical variation of dementia in New Zealand (NZ) from 2012-2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure (IDI)*
- September 2019: Poster presentation at British Society of Population Studies. *Prevalence and geographical variation of dementia in New Zealand (NZ) from 2012-2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure (IDI)*
- August 2019: Poster presentation at GeoMedicine conference. *Prevalence and geographical variation of dementia in New Zealand (NZ) from 2012-2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure (IDI)*
- June 2019: Alzheimer Scotland Conference. *Prevalence and geographical variation of dementia in New Zealand (NZ) from 2012-2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure (IDI)*
- April 2019: Scottish British Geriatrics Society Spring Meeting. *Prevalence and geographical variation of dementia in New Zealand (NZ) from 2012-2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure (IDI)*

Additional related key events:

- 2023: Contributed to the 2023 Sudlow Review on Big Data in the UK as

part of the UK Ageing Data Research Collaboration (ADRC)

- 2023: Co-founding member of EUGMS Special Interest Group on Ageing Data Research
- 2020: Co-founding member of the UK Ageing Data Research Collaboration linked with the British Geriatrics Society
- 2020: Part of an expert clinical academic geriatric medicine Delphi panel reviewing data to for a minimum geriatric medicine research dataset (paper published)
- 2017: Alzheimer Association International Conference (AAIC) London 2017 Travel Fellowship awarded to attend and present research as poster
- 2016: Medical Research Council PhD studentship Travel Grant awarded for travel to New Zealand to set up collaboration looking at dementia prevalence

Chapter 1: Introduction

1.1 Chapter introduction

This doctoral thesis explores how dementia risk varies with location – specifically between rural and urban areas - with a particular focus on how large routinely collected health data ('big data') can be used for investigating this variation, while concurrently discussing the challenges associated with this approach.

This chapter provides the rationale underlying this research, in addition to the aims and research questions this thesis will investigate, before outlining the structure of the remaining chapters. This introductory chapter will give:

1. An overview of dementia in terms of definition, prevalence, and diagnosis (and the complexities with this);
2. A brief overview of predictors of dementia, including life-course epidemiological factors;
3. An overview of what is meant by 'big data';
4. An overview of what is meant by location/place: with discussion of geographical determinants of brain health, and consideration of the relationship between geography, ageing, and dementia risk.

1.2 Overview of Dementia

1.2.1 Dementia definition

Dementia is a loss of cognitive and memory skills, and also loss of executive function, in parallel with an impact on activities of daily living, (ADLs) [19, 20, 21, 22, 23]. The International Classification of Diseases (ICD) Version 11 formally describes dementia as: *"a syndrome due to disease of the brain, usually of a chronic or progressive nature, in which there is disturbance of multiple higher cortical functions, including memory, thinking, orientation, comprehension, calculation,*

learning capacity, language, and judgement" [24].

Dementia is an ‘umbrella term’ for different types (sub-types) of dementia [24, 25, 26, 27, 28, 29, 30, 31, 32, 19]. Sometimes dementia is referred to as ‘all-cause’ dementia, sometimes by the different sub-types of dementia. It is a syndrome caused by one or more of several diseases [25, 26, 27, 28, 29, 33, 34, 35, 36, 37]. Unlike other physical health conditions there are not clear biomarkers that can be readily used outside of a research setting. Therefore, there is complexity in defining dementia for several reasons: different methods for clinical diagnosis, significant variation between individuals, and lack of consistency of definition with key characteristics as the term ‘dementia’ can encompass different characteristics from different diseases.

The most common sub-type dementia is Alzheimer dementia (dementia due to Alzheimer disease) [38]. In 1906, Alois Alzheimer first described ‘plaques’ and ‘tangles’ on post-mortem evaluation of brain tissue [38, 39]. These plaques are now known to be comprised of amyloid beta peptide (amyloid- β), and intracellular protein tau comprising the tangles [38, 39]. Later in 1910, Emil Kraepelin first coined the term ‘Alzheimer dementia’ in relation to a clinical diagnosis, with Oskar Fischer and Solomon Carter Fuller [40] also discovering the connection between these histopathological findings and late-onset dementia to progress research in the field.

Alzheimer disease and Alzheimer dementia are often interchanged in clinical practice. However, Alzheimer disease is the process of developing plaques and tangles that may occur pathologically prior to the development of dementia. Dementia due to Alzheimer disease is also referred to as Alzheimer disease dementia to differentiate that an individual has dementia due to Alzheimer disease.

Alongside the discovery of the link between the pathological changes and clinical changes in Alzheimer dementia, these researchers also began to consider the terms

‘pre-senile’ and ‘senile’ dementia [38, 39]. For the purposes of this thesis, I focus on late onset dementia (previously referred to as ‘senile’ dementia) with onset at >65 years old.

Approximately 30 years ago, the amyloid cascade hypothesis was discovered as potential theory in the development of Alzheimer disease [38, 39]. Since then, it has become a focus to target possible treatment options with recent excitement of the prospect of disease modifying drugs targeting amyloid- β (Aducanumab, Lecanemab and Donanemab) [41, 38, 42]. The last drug approved for dementia treatment, prior to these, was memantine over 20 years ago in 2003 [38].

Since the discovery of the amyloid cascade hypothesis, it has become apparent that the association between dementia and amyloid- β accumulation is not linear [38, 43, 44]. Therefore, there are still many more questions than answers suggesting a more complex aetiology [38]. A recent review (2023) proposes that amyloid- β plays an important role in the definition of Alzheimer dementia, but that its aetiological role may be far less than the multiple pathogenic factors contributing to Alzheimer dementia [45].

Alongside the clinical and pathological changes highlighted above, there is some understanding of the role genetics play in dementia risk, most notably in young onset dementia (<65 years) [39, 46, 47, 48, 49, 50, 51, 52]. Key mutations in the *presenilin 1*, *2* and *amyloid precursor protein* (*APP*) genes are known contributors to early onset dementia [38, 39, 47, 53, 50, 51].

For late-onset dementia, the apolipoprotein E type 4 allele (*APOE- ϵ 4*) on chromosome 19 is linked with Alzheimer disease dementia [54, 55, 56, 57, 58]. Possession of an *APOE- ϵ 4* allele can bring forward the age of onset of Alzheimer Dementia considerably [39, 38, 55, 56, 57]. Genome-wide association studies (GWAS) have identified risk loci for dementia [59, 60, 61, 62].

Alongside genetic differences in the risks of dementia, there is growing interest in epigenetics influencing dementia and its risk factors [63, 64, 65, 66, 67, 68, 69]. Epigenetics refers functionally relevant mechanisms of changes to the genome that does not alter the underlying nucleotide sequence of the DNA but might alter the way a gene is expressed. Mechanistic examples of epigenetics include DNA methylation or histone modification that can result in changes in how the genes are expressed without altering the underlying DNA sequence [70, 71, 72, 73].

Epigenome-wide association studies (EWAS) in post-mortem Alzheimer disease dementia brains are highlighting areas of interest [63]. This has led to consideration of how environmental factors might therefore influence dementia risk, and impact on a person's genome across their life through epigenetic mechanisms. Studies are suggesting evidence that life experiences and environmental exposures may *"leave epigenetic marks on genes"* [63]. This will be explored further in Section 1.5.2 (Location: geography and brain health).

When the term dementia is not further specified, it is sometimes considered to be synonymous with Alzheimer dementia (or Alzheimer disease dementia, ADD), as it is the main cause of dementia [74, 20]. It is important to be aware of the other causes of dementia. The second commonest sub-type of dementia is vascular dementia [74, 75, 75]. Vascular dementia is a term used for dementia due to underlying vascular disease, i.e. damage to the brain due to underlying impairment or damage to the vasculature supplying the brain alongside other marks of neurodegeneration [74, 75].

It is now recognised that vascular diseases such as strokes can increase the risk of dementia up to two-fold [74]. As highlighted in the book chapter I co-authored during my PhD, there are many synonyms linked with the definition of 'vascular dementia' and 'mixed dementia' and there is complexity in the diagnosis [10]. Often the two main causes of dementia (Alzheimer disease dementia and Vascular

dementia) are not distinct, and frequently co-exist.

The underlying disease contributing to dementia can be helpful to understand by considering modifiable risk factors. Clinical diagnoses of the underlying sub-type of dementia can be difficult as a recent paper highlighted in those diagnosed with Alzheimer disease dementia there was under-reporting of cerebrovascular pathology at postmortem [76]. Understanding the pathophysiology of dementia, and different dementia sub-types, is important in understanding the disease process at a cellular level: it does not always correlate clinically.

Individuals can present with features of more than one sub-type of dementia and many dementias are coded as ‘unspecified’ dementia in medical records. In this thesis, I will consider dementia largely in its broadest term of ‘all cause’ dementia because it is recognised that coding of the sub-types of dementia are not always reliable or reported [77, 21], and over 50% of people with dementia have more than one brain pathology which could cause dementia [78].

1.2.2 Dementia prevalence

Dementia is a growing worldwide major public health concern. It is predicted to affect 152.8 million people worldwide in 2050 [79]. Global prevalence of dementia varies between countries [80, 81]. Low and Middle-Income Countries (L-MICs) worldwide are due to see the greatest increase in incidence and prevalence in dementia as life expectancy rises [82, 22, 83]. In High Income Countries (HICs), there are mixed opinions on the future predictions for dementia prevalence as it has been suggested dementia incidence has decreased [84, 85, 86, 87, 88, 89, 90].

Some studies have suggested a reduction in dementia incidence and prevalence, particularly for vascular dementia, and suggestions have included that improved management of stroke disease and cardiovascular risk factors (primary and secondary prevention) have reduced the risks for vascular dementia, alongside improvements in education [89, 91, 90, 91, 84]. Other studies have shown no reduc-

tion in predicted future dementia estimates [92].

Age remains the greatest risk factor for late onset dementia [93]. Forecasting predictions of dementia is complex. The Global Burden of Diseases (GBD) collaborators use three risk factors for dementia (high body-mass index, BMI, high fasting plasma glucose, and smoking) [94].

With an ageing population, dementia rates will likely rise in Scotland. In the UK, approximately one in fourteen people have dementia over the age of 65, and one in six people have dementia from the age of 80 [95]. In Scotland, dementia has been a key area of Scottish Government policy [96]. In theory it is estimated that approximately 45% of all dementia cases could be prevented or delayed at a population level if specific risk factors were eliminated or reduced [97]. For this reason, many countries have developed strategies to focus on dementia. The World Health Organisation (WHO) document in 2017 ‘Global Action Plan on the Public Health Response to Dementia 2017-2025’ highlighted the importance of developing effective, low cost strategies for reducing and delaying dementia onset that were contextually appropriate [98].

To accurately estimate dementia prevalence, and provide clinical care and services based on the numbers of people with a disease, it is important to have country level data that reflects the numbers of people living with dementia. Most countries are unable to accurately calculate the number of people living with dementia, and must use predictions and estimations based on different modelling. Scotland has moved from using EURODEM models (upper prevalence rates) and The Dementia UK Report (lower prevalence rates) to EUROCODE in the official estimates of dementia prevalence [99, 100, 101, 102, 103]. EURODEM consists of a systematic meta-analysis of twelve European studies to generate dementia estimates [103, 99].

EUROCODE was published by Alzheimer Europe in 2009 and consists of the

results from a group of population studies across Europe to calculate prevalence rates [100]. Alzheimer Europe updated these studies in 2019 following the same methodology as set out by the EUROCODE project [77]. Internationally, the Global Burden of Diseases (GBD) Dementia Forecasting Collaborators within the GBD Group produces estimates and predictions of dementia along with other diseases [79, 104, 105], and Alzheimer’s Disease International produced their recent report with worldwide dementia estimates [106].

Whilst EURODEM was a milestone in the epidemiology of dementia in Europe, methodologies varied in the different studies used for EURODEM predictions which was recognised as a limitation [107]. It has been suggested that there are arguments for the validity of these estimates (EURODEM and The Dementia UK Report) in Scotland [99]. In general this concern of validity could be applied to any prevalence estimates, including EUROCODE. It is unclear how accurate modelling dementia estimates in certain countries relate to other countries, particularly if the data used may not be representative of the people within the other country.

Furthermore, it is important to stress that ‘dementia prevalence’ estimates of those diagnosed with dementia do not represent the true picture of dementia prevalence. This is known as a ‘diagnostic gap’ [108, 109, 110]. This reflects the estimated gap between those diagnosed and those living with dementia but remaining undiagnosed in the population [109, 110]. Given the proportion of dementia diagnoses represents the ‘tip of the iceberg’ in terms of dementia prevalence, there are more individuals likely living with dementia without a formal diagnosis as demonstrated in Figure 1.1.

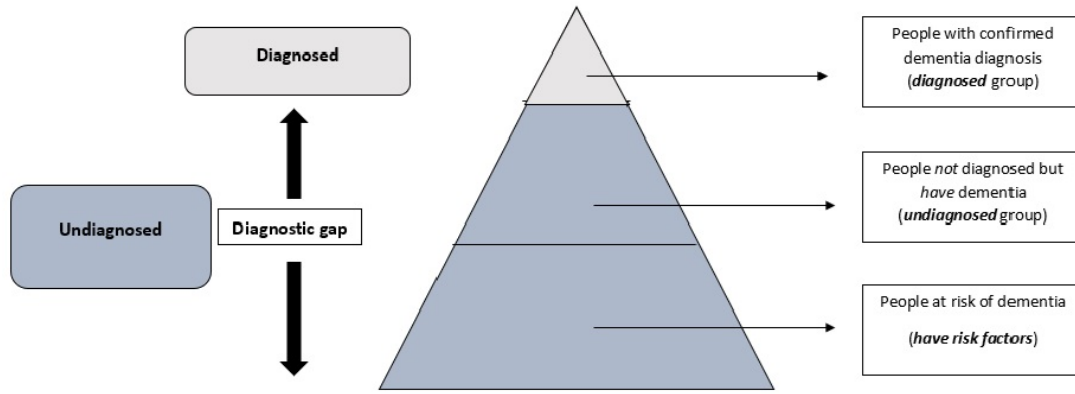


Figure 1.1: Explanation of the diagnostic gap via the ‘tip of the iceberg’ analogy

1.2.3 Dementia diagnosis

Dementia can be diagnosed through the presentation of clinical symptoms, radiological or histological signs [111, 112, 113, 29]. However, the diagnosis, as highlighted above in Section 1.2.1(Dementia definition), is not always clear cut. There are a number of clinical diagnostic criteria and tools used in the diagnosis of dementia [114, 115, 116, 117, 118, 26, 23]. Radiological imaging (in the form of different brain scans such as: CT, MRI, PET or SPECT scans) can highlight differences dependent on the sub-type of dementia, and stage in the dementia process [111, 112, 119, 120, 121, 122].

Diagnostic tests can also vary between a research or clinical setting. Other than imaging, biomarkers are not routinely used in clinical practice but may be used in research, or specialised clinics [123, 124, 125]. Evidence suggests that not everyone living with dementia seeks a clinical diagnosis, therefore there are likely many more people with biological signs of dementia who are still to be clinically diagnosed [126]. The reasons for seeking, or not seeking, a diagnosis of dementia are likely multi-factorial [127, 128, 129, 130, 131].

A clinical diagnosis of dementia is based on a clinical history, examination and assessment of a number of parameters including cognitive and behavioural/functional

factors [109, 132]. A summary of some of the key cognitive assessments used in clinical practice to diagnose dementia are listed in Table 1.1 [116, 133, 117, 118, 134, 135]. In the UK these are normally undertaken in a ‘memory’ clinic in secondary care (under neurology/psychiatry or geriatric services) although some of the initial investigations and assessments will be done in primary care [109, 132].

Whilst treatment options are limited, treatment is not the only outcome from a diagnosis of dementia. Other management includes supportive post-diagnostic support (for the individual and family) [136, 137, 138, 139], understanding behaviour that some people might find challenging, developing individualised care plans and also future planning [140].

Cognitive assessment	Began	Summary
Mini-Mental State Examination (MMSE) [116]	1975	Clinical and research settings. Assesses 5 cognitive domains: orientation, registration, attention and calculation, recall, and language. Screening tool for Mild Cognitive Impairment (MCI) and dementia.
10/66 Protocol [134]	1999	Clinical and research settings (used in certain clinical settings/groups including L-MICs or trialled in Indigenous populations within HICs). Developed for multi-cultural or low education settings. Developed as a language and culture-fair instrument for non-English speaking areas.
Montreal Cognitive Assessment (MOCA) [133]	2000	Clinical and research settings. High sensitivity for early changes/MCI. Assesses short term memory, visuo-spatial abilities, executive functions, attention, concentration, working memory, language and orientation. Multiple versions.
Addenbrooke's Cognitive Examination (ACE) [118, 141]	2000	Clinical and research settings. For MCI, dementia and dementia sub-types. Assesses 6 cognitive domains: orientation, attention, memory, verbal fluency, language and visuo-spatial. Multiple revisions.
General Practitioner Assessment of Cognition (GP-COG) [117, 142]	2002	Clinical and research settings. Screening tool for cognitive impairment. Quicker than, and popular alternative to, MMSE. Multi-cultural.

Table 1.1: Summary of a selection of key cognitive assessment tools

There are different classification systems or criteria to follow for diagnosing dementia sub-types, which have been revised and/or improved over time [115, 143, 144, 145]. The current key classifications are: ICD-10 and ICD-11 (all diseases) [145, 21], NINDS-AIREN (Vascular dementia) [146], NINCDS-ADRD Alzheimer criteria [25], National Institute of Ageing (NIA) for Alzheimer disease [126, 26],

and DSM-V (all psychiatric diseases) [147, 148]. The ICD-11 was officially endorsed by WHO members in 2019 following a decade of specialists preparing it [21].

This thesis uses the earlier ICD-10 version (and also ICD-9) to define dementia, due to the delays between endorsing the new version (ICD-11) and its implementation in the clinical setting. It is not yet introduced for all coding in Scotland, but is being used by mental health. The linked data in the National Safe Haven (NSH) have not mapped ICD-10 codes to ICD-11 codes yet. Previous ICD codes are converted to the latest ICD version, but this can take time for historical records. The Diagnostic and Statistical Manual for Mental Disorders (DSM) is in its fifth revision (DSM-V) as of 2013 [148]. In some countries DSM-IV is the principal book for diagnosing mental health diseases. In other countries, DSM-IV is used in conjunction with other guidelines or documents. Overall, the DSM puts more focus on clear diagnostic criteria whereas ICD allows more clinical judgement in the coding [148, 21].

There have been several revisions of criteria for Alzheimer disease and mild cognitive impairment over the last twelve years. A key theme is that the pathophysiological process takes years and decades prior to clinical symptoms being fully evident for a full clinical diagnosis of dementia, and a more recent focus on defining the disease biologically rather than a syndrome [114, 26, 149, 150]. Labels of *‘possible’*, *‘probable’*, and *‘definite’* Alzheimer disease dementia are often used [114, 26]. The first two are used in clinical settings and the third is used in research settings [114, 26]. *‘Probable’* Alzheimer disease dementia is the maximum level of certainty without pathological evidence [26]. *‘Definite’* Alzheimer disease dementia needs supporting pathological evidence [114, 26].

Separately, the National Institute of Ageing (NIA) criteria for Alzheimer disease highlighted a number of areas of the original criteria that required revisions [26]. One such area was the heterogeneity in the *‘Possible’* Alzheimer disease dementia

category, in which many patients would subsequently be diagnosed as having Mild Cognitive Impairment (MCI) [26]. This highlights that there are several stages in the development of dementia, particularly Alzheimer disease dementia (pre-clinical, Mild Cognitive Impairment due to Alzheimer disease, and Dementia due to Alzheimer disease) [114, 26].

There are also terms used in general, not specific to Alzheimer disease or Alzheimer disease dementia, that are used for the stages in the development of all-cause dementia (pre-clinical, prodromal, and dementia stages) [26, 114]. The ‘pre-clinical’ stage is important in relation to research and considering at what period in a person’s life might modifiable risk factors or treatment options be best targeted or how an understanding of the disease process of dementia starts before we see clinical signs as clinicians [114, 26]. This is outlined in Figure 1.2 [151].

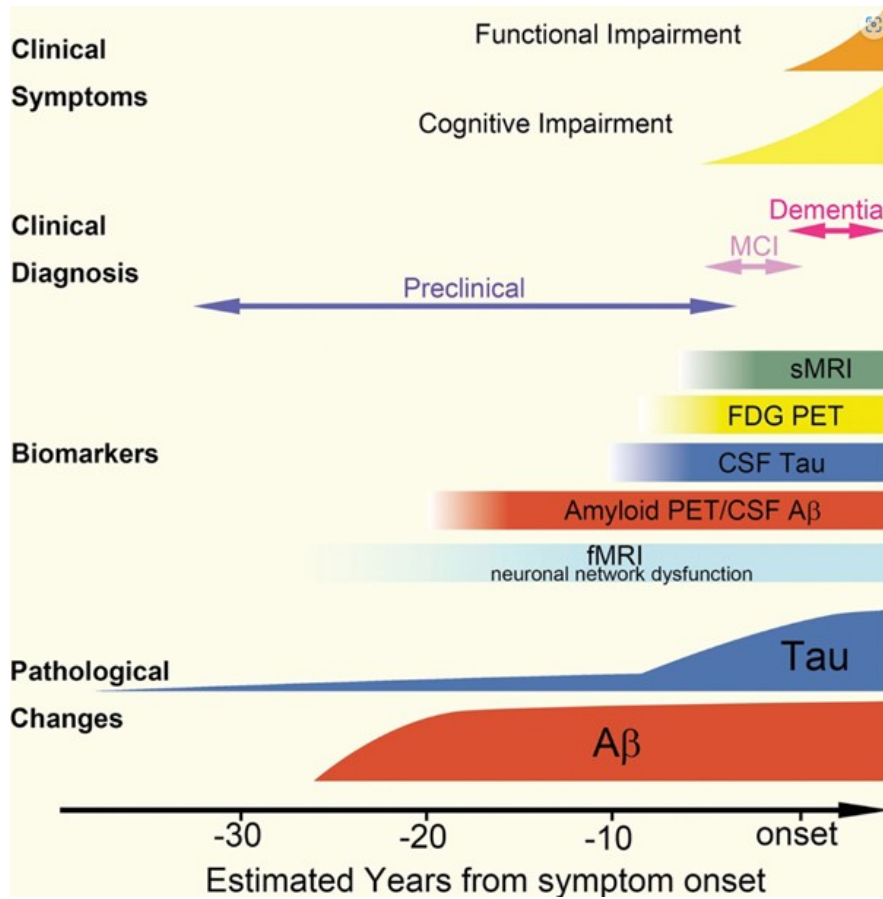


Figure 1.2: Chronological relationship between pathology, clinical symptoms and biomarkers [151]

Source: Reproduced from *Therapeutic strategies for tau mediated neurodegeneration* in *Journal of Neurology, Neurosurgery & Psychiatry*, Authors: Yasumasa Yoshiyama, Virginia M Y Lee and John Q Trojanowski, Vol 84, Page 785, Copyright 2013 with permission from BMJ Publishing Group Ltd.

In the context of this thesis, it is helpful to highlight this complexity as I use routinely collected data for ascertaining dementia (largely through ICD coding of dementia or sub-types of dementia). I am not able to access the clinical/radiological or pathological diagnosis process that precedes the ICD coding of dementia, i.e., the *quality* of the diagnosis. This is a limitation in ‘big data’ research using coding for dementia ascertainment in routinely collected health data. This overview has shown that a dementia diagnosis is not always clear cut. In [Chapter 3: Using big data for dementia research](#) (Section 3.8 Routinely collected data: dementia)

I discuss that having a diagnostic code for dementia has been shown to have a positive predictive value that an individual *has* dementia [152]. The numbers of people who may have dementia but are yet to have a diagnosis is more challenging to demonstrate using routinely collected health data for epidemiological research.

1.3 Overview of ‘big data’

In [Chapter 3: Using big data for dementia research](#) I discuss using big data for dementia research, with a focus on secondary use of routinely collected healthcare data, data linkage and the health data landscape in the UK. In this section of this introductory chapter, I first introduce what is meant by the term ‘big data’ and also discuss what are ‘routinely collected health data.’

1.3.1 Defining ‘big data’ in healthcare

The term ‘big data’ has become an increasingly popular term within healthcare [153, 154, 155, 156, 157, 158, 159]. Defining what constitutes ‘big data’ is complex, often without a uniform definition [160, 161, 159].

Within healthcare as clinicians, along with all services users, we are constantly creating big data. A patient’s journey from ‘cradle to grave’ is captured at each contact with the National Health Service (NHS) as part of the routinely collected data within their Electronic Health Records (EHRs). These are added to - often on a daily basis - as part of their clinical care: clinical notes recorded by members of the multidisciplinary team; observations, such as temperature, pulse, blood pressure; blood test results; and imaging results or raw images; medication prescriptions. Data from healthcare can be linked with data from other domains including social care or geographical information to gain insights into the causes and consequences of disease.

‘Big data’ is not a new concept. In the 1800s, William Farr (an epidemiologist) realised the potential to analyse routinely collected data [162]. By collecting infor-

mation on the causes of death he was able to highlight variations geographically and initiate public health changes to improve lives. This was, in part, also driven by the work of Dr John Snow (a physician) collecting and analysing healthcare and geographical data to understand the role that the ‘Broad Street Pump’ had in the cholera epidemic in London in 1853-1854, ahead of the discovery of the pathogen [163]. This is an excellent – and legendary – example of using collected data to investigate a research question [163].

Whilst the consensus in what ‘big data’ means varies in different settings, it is recognised that key components are the data’s size, scale and volume or the ability to link large datasets [164, 160, 161, 159]. ‘Big data’ can be described by the "7V’s" it encompasses (see Figure 1.3) [165]. I highlighted this in the published book chapter *"Using big data in primary care research"* I co-authored and include in 7.7 [4]. The number of "V’s" have been extended in recent years as the use of ‘big data’ increased, from Garner’s "3V’s" (Volume, Velocity and Variety) [164] highlighted in my peer-reviewed publication [2] in Section 3.2 (Peer-reviewed published manuscript) from 2017.

Volume – the scale and size of the data
Variety – different types of data sources available (images, text etc.)
Velocity – the speed of data generation
Veracity – the accuracy of the data
Value – new insights and understanding of value to the organisation and patients
Variability – consistency of the data (and reproducibility of results)
Visualisation – ability to communicate patterns of data through novel techniques

Figure 1.3: The 7 V’s of ‘big data’: text extracted from *The 7 Vs of Big Data – what are they and why are they so important?* [165]

1.3.2 Defining ‘routinely collected data’

For the purposes of this thesis relating to healthcare research, when I refer to ‘routinely collected data’ I refer to health and administrative routinely collected data used by healthcare services or used to understand health and disease. Unless otherwise clearly stated, I am not referring to routinely collected data which numerous companies and organisations collect on a daily basis e.g. google searches or fitbit data downloads.

The term ‘routinely collected data’ is often used interchangeably in healthcare with ‘big data.’ Routinely collected healthcare data consists of information collected from a person’s interaction with healthcare services. They include a number of healthcare observations, interventions or outcomes as well as individual level information on the patient and their health and diseases. Within healthcare, as most healthcare is largely digitised in High Income Countries (HICs), the phrase ‘routinely collected health data’ is often synonymous with the term EHRs [166]. EHRs record a patient’s interaction, appointments, diagnoses and background personal information relevant for a healthcare setting or a healthcare provider to be able to provide healthcare [166].

Routinely collected data also varies between areas, and countries, depending on how much of this information is digitised. In the UK, as the digitisation of NHS records progresses, the majority of key interactions between patients and providers are recorded within datasets. In relation to dementia care, an individual may attend their GP with concerns about their memory, which would initiate a digital record of this interaction including any blood tests, with potential further onward investigations/referral to secondary care memory clinic services and associated blood tests, and imaging.

Outpatient or inpatient hospital attendances are digitally recorded in discharge letters and in prescribing datasets if medications are prescribed. Each interaction

or investigation, or coding of a medical diagnosis, cumulatively builds up a picture over time of that individual's health and their interaction with services linked together by an individual identifier such as CHI-number in Scotland which will be explained in Chapter 3 Section 3.5.1 (Unique identifiers). This can prove a valuable resource for research. The recent (2022) document by the Health Foundation in the UK entitled 'How better use of data can help address key challenges facing the NHS' [167] highlights how: *"routine NHS data can directly support care, drive service improvement and enable research and innovation"*.

Routinely collected data and 'digital health' are expanding within the UK and internationally. The World Health Organisation (WHO) recommendation is for *"person-centred digitalisation"* [168]. Their report, 'Global strategy on digital health 2020-2025' [168], outlines a worldwide approach to strengthen health systems through digital health technologies and create a shared understanding of the *"importance of digital health solutions, and an approach towards creating an interoperable digital health ecosystem"* that can be used by health providers, patients, public health authorities, universities and research institutions [168].

In Chapter 3: Using big data for dementia research, I discuss how routinely collected health data can be repurposed with a secondary use for research purposes in addition to their intended clinical use. Whilst these datasets were traditionally not originally designed for research purposes, they do allow this 'unique window' into the interactions of an individual using healthcare services. This then provides a potential, and valuable, 'real world' dataset which can be used for research, with appropriate permissions and approvals as I will further discuss, with a focus on dementia research use, in Chapter 3.

1.4 Overview of dementia epidemiology

Epidemiology is the study of the distribution of diseases, and factors contributing to/predicting this. This sub-section discussing prediction uses epidemiology to

discuss the prediction of dementia.

We still do not have an in-depth understanding of the epidemiology of dementia. Some aspects can be explained epidemiologically, but much remains unknown. Who develops dementia and why; and why one person follows a more ‘predicted’ trajectory with dementia progression and another does not is unknown. Whilst dementia sub-types, accumulation of risk factors, and subsequent multiple long-term conditions, in addition to pre-morbid and early life baselines, influence this trajectory, there are still far more questions than answers in understanding the whole picture.

Epidemiology’s ability to inform causality is challenging, and is dependent on the quality of the data used and ability to access data relevant to the question. For a condition such as late onset dementia where prevalence is highest in the >65-80 year group onwards, there may be many points across an individuals life-course that influence their risk. Birth cohorts are traditionally very valuable in epidemiological research [[169](#), [170](#), [171](#), [172](#), [173](#), [174](#), [175](#), [176](#), [177](#), [178](#)]. Additionally, traditional epidemiological birth cohort studies are being linked with routinely collected health data for research [[179](#), [180](#), [181](#)].

However, their use in studying dementia is complex as a recent birth cohort will not be of value for a disease presenting in later life for decades. An established birth cohort with over 70 years of data is the ideal situation for dementia epidemiological research, particularly investigating the life-course epidemiological perspectives of dementia. There are not many birth cohorts to do this type of research, and some have limitations due to cohort effects, missing data, and bias. In this section I discuss further epidemiology in relation to dementia, and then focus on discussing the life-course epidemiological perspective of dementia in more detail in the subsequent section.

It is proposed that by addressing preventative or potentially modifiable risk fac-

tors, that dementia may be prevented or delayed at a population level in around 45% of cases [182], although this has not been fully tested in intervention studies (i.e. whether individuals will necessarily fully change their behaviour influencing their risk). This awareness, combined with a greater understanding of there being a ‘pre-clinical’ phase of dementia as discussed in Section 1.2.3 (Dementia diagnosis) focuses attention on making changes years before the clinical signs of dementia are visible (and a dementia diagnosis made).

Figure 1.4 highlights this. It shows the time lag from when pre-clinical signs of dementia (evident in biomarker changes) are present and later the trajectory into the clinical disease stages of Mild Cognitive Impairment (MCI) and lastly dementia [183]. A referral for short term memory loss might be the key symptom noticeable by a person, or their relative, but underlying changes have likely been occurring for many years.

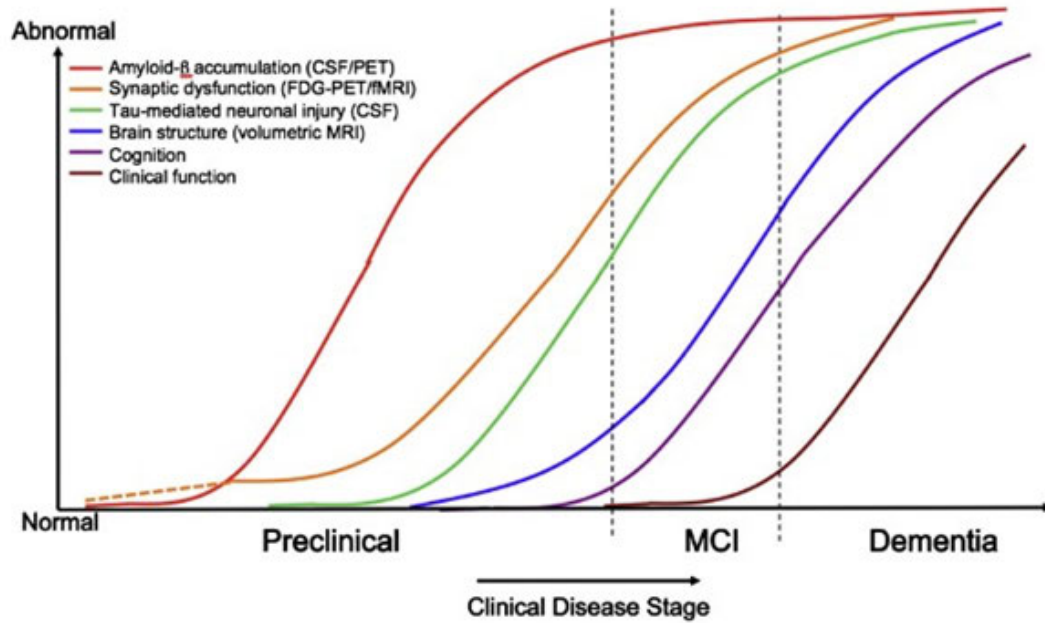


Figure 1.4: Hypothetical model of biomarkers for Alzheimer dementia from [183], which is based on the original from [184]

Source: Reproduced from *Alzheimer's and Dementia*, Authors: Reisa A. Sperling, Paul S. Aisen, Laurel A. Beckett et al., Title: Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Vol 7, Page 285, Copyright 2011. Used with permission from John Wiley and Sons, and Copyright Clearance Center.

As discussed earlier in this chapter, dementia encompasses many sub-types and therefore different aetiologies and therefore different approaches to modifiable risk factors, although the common themes of prevention in the main sub-types of dementia remain. Therefore, there is unlikely ever to be a drug that is suitable and effective for all with dementia and a person-centred approach looking at individual risk factors and modifications may produce the greatest impactful change. Additionally, understanding our modifiable risk factors might allow for more appropriate targeting of future dementia treatment (including medications). One approach to modifying risk factors for dementia in research is by focusing on 'healthy brain ageing' and/or 'brain health' [185, 186, 187, 188, 189, 190].

There are two analytical approaches to risk factors and dementia: 1) observational

(case-control, cohort, cross-sectional) and 2) experimental (clinical trials). For either approach, the risk factors for dementia can alter across a person’s life-course both in terms of what those risks are and whether certain points in a person’s life are more ‘critical’ or ‘sensitive’ to those risk factors.

1.4.1 Life-course epidemiology in relation to dementia

Life-course epidemiology typically involves several pathways: biological, behavioural and psychological over a person’s life. Conceptual models in life-course epidemiology can conceptualise and illustrate potential pathways between a later life outcome in adult disease and early life exposures [191]. This can allow a researcher to investigate childhood factors influencing later life disease but also potential pathways with intermediaries or confounding factors [191].

Traditionally, there has been a need to include a multivariate model of exposures/variables with the temporal relationship between these variables. However, where an individual lives and the ‘spatial’ (geographical) environment may change over time. There can be interactions between the individual and the geographical location over time and the variable of interest in terms of exposure may be influenced by the location and any changes in this geographically or environmentally. A spatial element needs also considered in how it might be influencing exposures/variables and that this is also not static over time. There is also a need to consider changes in spatial relationship, as part of the temporal-spatial relationship between variables. It is important to consider the inter-relationships between exposure variables and the individual. The way an individual interacts with their ‘location’ can vary with time, due to multiple factors, which then might influence the exposure variable over time.

For late onset dementia, life-course epidemiology is complex. As dementia prevalence increases with age, the life-course approach of considering multivariate models of exposures and the spatial-temporal relationship between them is over 60-80 years, or longer. Highlighting an association between different variables and de-

mentia risk is possible, but demonstrating a causal relationship can be more challenging due to the lengthy time period, and potential spatial variation.

There are many potential exposures across an individual’s life and their subsequent risk of dementia as well as understanding the importance of the timing of those exposures. It is not surprising that exposures may affect the risk of disease in more than one way. A previous editorial [191] proposed a simple classification of “*potential life course models of health*” (Table 1.2) [191].

Critical period model
with or without later life risk factors
with later life effect modifiers
Accumulation of risk
with independent and uncorrelated insults
with correlated insults
‘risk clustering’
‘chains of risk’ with additive or trigger effects

Table 1.2: Conceptual life-course models [191]

Source: International Journal of Epidemiology, Authors: Ben-Shlomo, Yoav; Kuh, Diana, Title: A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives, 2002. Used with permission under licence from Oxford University Press and Copyright Clearance Center.

Critical periods may include developmental periods, during which the mechanisms in biological subsystems are developed [192, 193, 191, 194]. For instance, this links with the developmental origins of adult disease hypothesis [192] where there is biological programming in early life and then a latency period before adult disease occurs but that the early prenatal life events have influence on this. Evidence suggests that the first few years of life are influential on later life [192, 193, 191, 195, 196, 194]. Another potentially relevant model is the life-course accumulation

of risk, whereby an individual accumulates risks across their life time [193, 191, 197]. This might be influenced by critical periods but also ‘sensitive periods’ in a person’s life. Sensitive periods may be more common throughout a person’s life and may also influence behavioural development.

Chronic non-communicable diseases often have multiple causes and risk factors, that contribute to their development, which are arguably linked in a causal pathway. There are broadly four types of factors involved in causation of a disease. These factors might be modifiable, and at certain key stages of a person’s life-course, influence their future risk of disease. Table 1.3 highlights the definitions of these four factors from Bonita et al. 2006 [193], and Figure 1.5 demonstrates how they may link towards risk in a non-communicable disease. I will later explore in this section what specific risk factors may be involved in developing dementia and at what stage in a person’s life they might be important.

Factors	Summary
Predisposing	Age, sex, specific genetic traits, or previous illness creating susceptibility to a disease agent
Enabling (disabling)	Low income, nutrition, housing and access to medical care increasing the risk of disease development. Or the reverse as protective factors
Precipitating	Exposure to a specific disease factor
Reinforcing	Repeated exposure, the environment or daily situation (e.g. occupation) that aggravates established disease

Table 1.3: Definitions of the four factors in causation [193]

These factors might be implicated in risk of a non-communicable disease. They can be split into social, environmental, lifestyle and physiological categories of different risk factors, although, it is difficult to segregate them rigidly. For instance it could be argued that psychological factors are minimised to "psychosocial" and "lifestyle." Furthermore, Figure 1.5 highlights one direction, whereas it could be

recognised that some risk factors (and population level public health policy) need a bi-directional approach, depending on many influences across an individual's life-course. One benefit from this approach is to allow a multi-modal approach to interventions from population level public health policy to individual primary and secondary prevention strategies (personalised care). This is very well illustrated in Figure 1.5 from Bonita et al. 2006 [193].

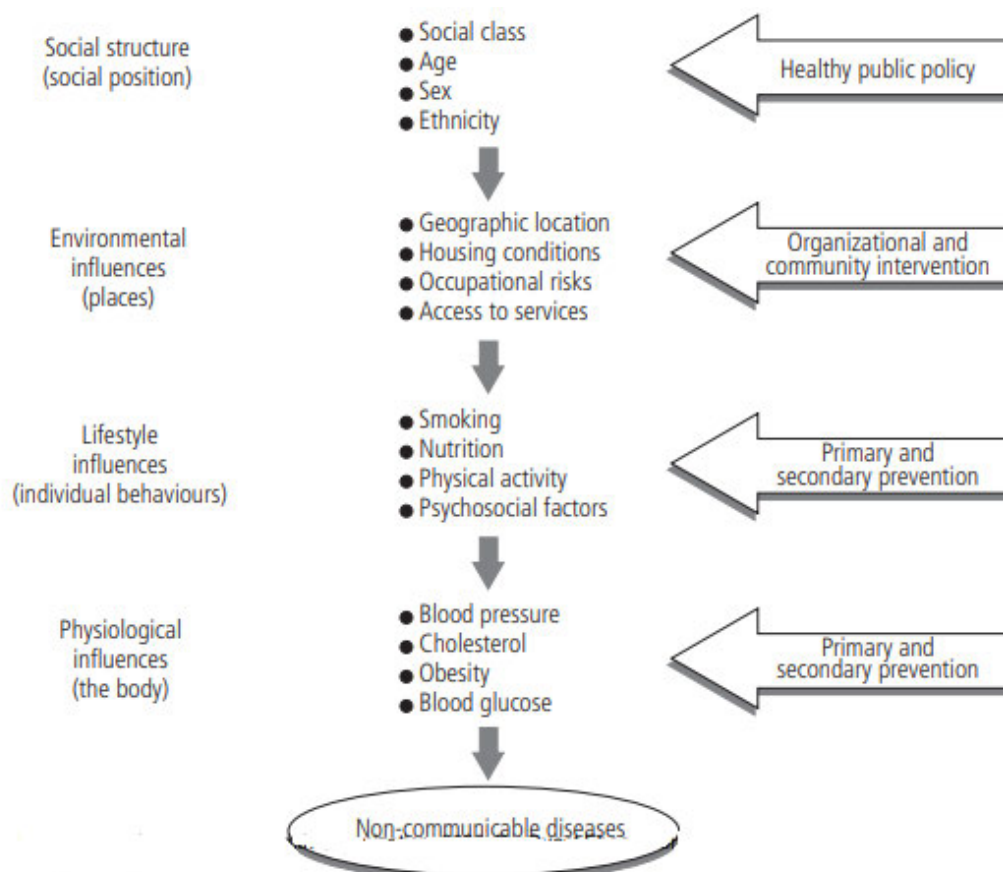


Figure 1.5: Common risk factors in non-communicable diseases [193]

*Source: World Health Organisation. Authors: Bonita, R., et al. Book Title: Basic Epidemiology, World Health Organization, 2006. Used with permission under CC BY-NC-SA 3.0 IGO licence for non-commercial educational purposes:
<https://creativecommons.org/licenses/by-nc-sa/3.0/igo/>*

When considering dementia as the non-communicable disease in Figure 1.5, it is important that individual risk factors in each section are not seen in isolation, thus

demonstrating why life-course epidemiology is important for dementia research [193, 191]. An individual’s risk factors across their life-course need to be considered, with the temporal linkage between these risk factors over an individual’s life. As discussed above there may be ‘critical’ periods for risk or ‘accumulation’ of risk and these are relevant to dementia life-course epidemiology [193, 191].

1.4.2 Risk factors for dementia

The Lancet Commission on Dementia Prevention, Intervention, and Care described a life-course model of dementia prevention in their three reports in 2017, 2020 and 2024 [182, 198, 97]. The 2020 Lancet Commission on Dementia report incorporated twelve population level modifiable risk factors identified from the literature [182]. The 2024 revised report highlighted three further risk factors: Traumatic Brain Injury (TBI), alcohol excess, and air pollution that had not been included in the earlier 2017 or 2020 Lancet Commission on Dementia Prevention reports [182, 198]. The latest 2024 report has highlighted an additional further two population level potentially modifiable risk factors (visual loss, and high LDL or cholesterol) [97]. These risk factors are highlighted in Table 1.4, and Figure 1.6 [97].

The Lancet Commission on Dementia prevention reports [182, 97, 198] were unique because they took a life-course approach to dementia prevention and calculated the contribution to risk of dementia at different points across an individual’s life span. These are represented as Population Attributable Fractions (PAFs). PAFs are estimated fractions of all cases of dementia that would not have developed if there had been no specific exposure [199]. From this it can then be proposed that there would be approximately 45% fewer cases of dementia in the 2024 Report [97] (Table 1.4) if these potentially modifiable risk factors were removed, though this has not been tested in trials.

Early life	PAF	Mid-life	PAF	Late life	PAF
Less Education	5%	Hearing loss	7%	Social isolation	5%
		High LDL or Cholesterol	7%	Air pollution	3%
		TBI	3%	Visual loss	2%
		Depression	3%		
		Smoking	2%		
		Hypertension	2%		
		Physical inactivity	2%		
		Diabetes	2%		
		Obesity	1%		
		Excess alcohol	1%		

Table 1.4: Population attributable fractions (PAFs) and dementia risk extracted from the text of The Lancet Commission for Dementia Prevention 2024 report [97]

Table 1.4 and Figure 1.6 highlight that in isolation the individual risk factors alone do not contribute substantially to the percentage risk reduction but are correlated [97]. Therefore, a holistic approach to addressing an individual’s risk of dementia and their potential for multiple co-morbidities could prove far greater risk reduction overall.

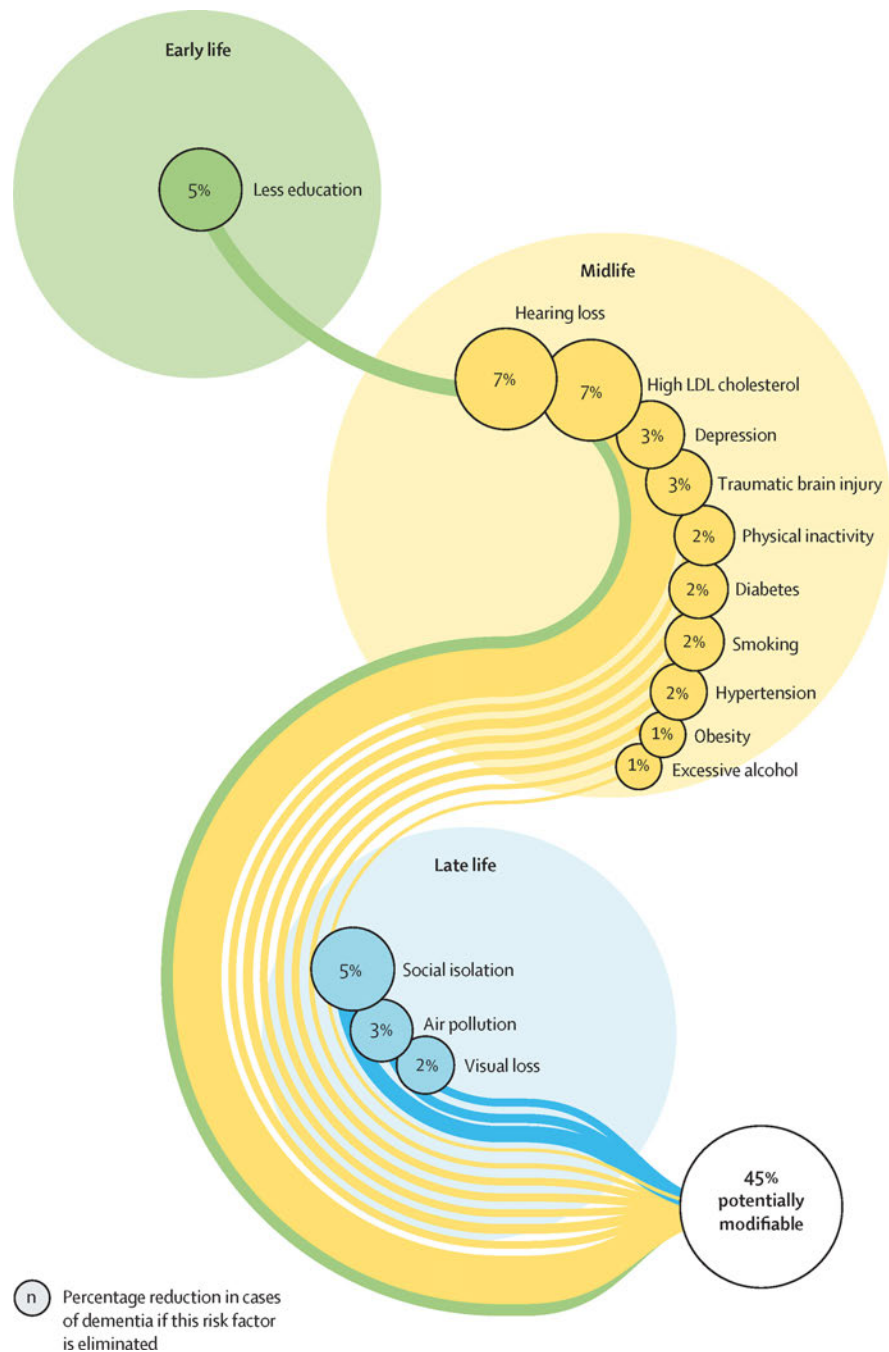


Figure 1.6: Population level modifiable risk factors contributing to dementia in a life-course model [97]

Source: Reprinted from *The Lancet*, Vol. 404, Authors: Gill Livingston, Jonathan Huntley, Kathy Y Liu et al., Title: *Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission*, Page No. 593, Copyright (2024), with permission from Elsevier and Copyright Clearance Center.

In the 2017 Lancet Commission dementia prevention report, the potentially non-modifiable risk factors contributed 65% which included genetic risk (APOE ϵ 4, 7%) and other factors [198]. In the 2020 Lancet Commission dementia prevention report, this changed to ‘risk unknown’ in 60% [182]. The additional three population level potentially modifiable risk factors introduced in this second (2020) report contributed a further 5% potential risk reduction in dementia [182]. In the recent 2024 version of the Lancet Commission dementia prevention report an additional two population level potentially modifiable risk factors (visual loss and high LDL cholesterol) were added [97]. This resulted in a further 5% potential risk reduction of Population Attributable Fractions (PAFs) in dementia, and overall a reduction by 45% of potentially modifiable risk factors for dementia [97]. Furthermore, most of these (ten) were in midlife in the latest report [97] which highlights this as a key area for future research in dementia prevention.

This section demonstrates that there are clearly risk factors for dementia across a life-course, some of which have clear and consistent relationships across different research papers, while others are inconclusive. Overall, the risk factors which are potentially modifiable may contribute to a 45% risk reduction in dementia [97]. However, the causal pathway of some of these risk factors is still unclear, and further research is needed as well as understanding the temporal change in the risk factors across a person’s life.

1.4.3 Brain health

Following the first two reports from the Lancet Commission’s work on dementia prevention [182, 198], there has been a growing focus on ‘brain health’ in public health approaches and research. The positive language of moving to ‘health’ rather than a ‘disease state’ can help individuals feel more control over modifiable risks. As defined by the WHO, brain health refers to: *“The state of brain functioning across cognitive, sensory, social-emotional, behavioural and motor domains, allowing a person to realize their full potential over the life course, irrespective of the presence or absence of disorders”* [200].

In 2022, WHO produced a position paper on ‘Optimizing brain health across the life course’, which discusses a framework for brain health optimization and highlights the determinants related to brain health [201]. These are categorised into five broad groups: 1) Physical health, 2) Healthy environments, 3) Safety and security, 4) Learning and social connection and 5) Access to quality services [201]. The document also stresses the need for a holistic person-centred approach and future direction for research, policy and data [201].

Whilst this document looks at brain health overall, complementing the associated WHO document on ‘*Intersectoral global action plan on epilepsy and other neurological disorders 2022–2031*’ [202], it also highlights dementia as a common neurological condition [201]. Given the known risk of early life influencing the risk of dementia in later life, Figure 1.7 demonstrates the importance of brain health when considering dementia risk through the life-course [201].

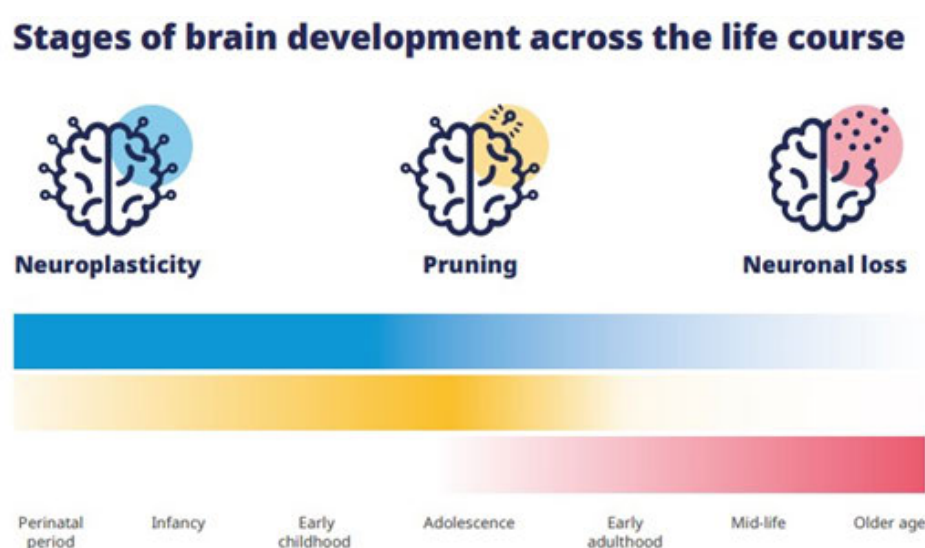


Figure 1.7: Stages of brain development across the life-course [201]

Source: World Health Organisation (WHO), Author: WHO, Title: Optimizing brain health across the life course: WHO position paper, Copyright 2022. Used with permission under Creative Commons Licence CC BY-NC-SA 3.0 IGO.

The Global Council on Brain Health (GCBH) is an independent collaboration of a large number of experts to educate the public and promote older people’s cognitive well-being [83, 203]. In essence, their approach is to translate the risk factors highlighted in the section above into a series of attainable positive healthy behaviours to sustain brain health. To do this they have created ‘Six Pillars of Brain Health’ [204] incorporating the GCBH’s analysis of modifiable risk factors. Table 1.5 highlights the six pillars and corresponding information for each of the pillars [204] that contribute to brain health, and have been produced in a number of reports including the Lancet Commission reports [83, 182, 203].

Pillar	Recommendation
Be Social	Keeping in touch with friends/family and preventing isolation.
Engage Your Brain	Finding ways to stimulate thinking, including new interests and hobbies.
Manage Stress	Adopting a stable daily schedule, including relaxation.
Ongoing Exercise	Aiming for 2.5 hours of moderate physical activity weekly and movement throughout each day.
Restorative Sleep	Having 7-8 hours of restful sleep daily.
Eat Right	A nutritious, heart-healthy diet (fish/vegetables/fruit).

Table 1.5: The Six Pillars of Brain Health extracted “How to Sustain Brain Healthy Behaviors: Applying Lessons of Public Health and Science to Drive Change.” [204]

The GCBH highlights an important population health component, which is to translate life-course risk factors for dementia into attainable and relevant health goals and therefore help modify potential risk factors for dementia [83]. Next, I consider early and mid-life risk factors for dementia. Since we know that pre-clinical changes in dementia are seen long before clinical signs of dementia [205,

[183, 206], these time periods in a person’s life might be very important in reducing the risk of dementia.

1.4.4 Importance of early/mid-life dementia risk factors

Some evidence suggests that there may be an accumulation of risk or ‘chains of risk’ across a person’s life time increasing their likelihood of developing dementia [191]. This might start in childhood [192, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220]. The Developmental Origins of Health and Disease theory suggests that the risk of adult onset non-communicable diseases is associated with “*adverse environmental conditions during embryonic and/or foetal development*” [63, 192].

It is known that the first few years of life are influential on overall later life health. Another area of increasing interest in life-course approaches to dementia risk, is the mid-life period [182, 221, 222, 223, 224]. This is in part linked with the knowledge that in research pre-clinical changes for dementia occur ten to twenty years prior to dementia showing clinically [183, 26, 206, 225, 205]. In this sub-section I discuss these two life periods.

A recent (2019) systematic review [211] investigating early-life risk factors for dementia and cognitive impairment in later life found that education level, leg length, history of childhood head trauma, family-related factors and learning ability were associated with dementia and cognitive impairment risk later in life. Leg length (or height) is a proxy for several early life factors [226, 227, 228]. In the meta-analysis of this systematic review, three factors in childhood were associated with increased risk of dementia in later life: food deficiency, leg length and education attainment [211]. Meta-analysis of these cross-sectional retrospective studies [229, 230, 229] ($n = 4,011$) showed a higher risk of dementia and cognitive impairment in those who experienced food deficiency in infancy or childhood (OR= 2.05, 95%CI 1.22-3.44, $I^2 = 0\%$, $p = 0.88$) [211].

There were three studies [231, 232, 233] ($n = 2,727$) in the meta-analysis that reported shorter leg length had a 19% increased dementia risk with an odds ratio of 1.16 (95%CI, 1.07-1.27), and 1.18 (95%CI, 1.05-1.31) in the fixed and random effects models [229]. Finally, there were thirteen studies ($n = 31,532$) in the meta-analysis related to low education level and an 80% increase in dementia or cognitive impairment risk - relative risk of 1.80 (95%CI, 1.60-2.02) for both fixed and random effects models [211]. The Lancet Commission reports 2017, 2020 and 2024 for Dementia prevention [182, 198, 97] also highlighted low education level as a childhood risk factor for later onset dementia with a relative risk for dementia population attributable fraction of 1.6 (95%CI, 1.3-2.0) in the 2020 Report [182].

The Lancet Dementia Prevention Commission defined early life as <45 years old [182]. Interestingly, the Lancet Commission [182] highlighted that some studies suggested that whilst cognitive ability increased with education, it reached a plateau in late adolescence with little gains with education after 20 years of age, as maximum brain plasticity has been reached in late adolescence [234]. Additionally, the systematic review found two papers demonstrating that rural residence in childhood had an increased risk of dementia [211, 235, 236], which I discuss in more detail in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#).

As birth cohorts age, and also the ability to link them to routinely collected healthcare data improves, this will provide increasing data for the analysis of life-course factors, particularly childhood influences on dementia risk [169]. This may allow potentially modifiable risk factors for late onset dementia to become clearer across a person's life course.

‘Cognitive maintenance’ is a term to encompass the the ability to maintain cognitive skills as one ages [237]. During typical ageing there is a reduction in multiple cognitive skills [238, 237]. It is known that cognitive ageing differs in the trajectory between individuals [238, 237]. This could be influenced by many

factors. The concept of the brain reserve, cognitive reserve, compensation and maintenance have been of increased research interest in recent years to understand cognitive resilience that influences the trajectories between individuals and cognitive ageing [238, 237].

Mid-life risk factors may include an element of needing ‘cognitive maintenance’ through cognitive stimulating activities to reduce the risk of dementia [239, 240, 241]. It is difficult to separate adult cognitive activity from activities for those with more education. Several studies [242, 223], have tried to investigate this further. Many of the smaller studies show that cognitively stimulating activities, including knowing a second language, can influence cognitive maintenance. A large study also highlighted that those >65 years who engaged in cognitively stimulating activities had reduced risk of dementia with an odds ratio 0.7 (95%CI, 0.6-0.8) [243].

The Lancet Commission 2020 named several other mid-life potentially modifiable risk factors for dementia [182]. These include: hearing loss, traumatic brain injury (TBI), hypertension, and alcohol (they used the cut off of >21 units/week although the recommended ‘healthy’ consumption level is <20 units a week) and obesity [97]. In the latest Lancet Commission report on Dementia Prevention (2024) some later-life risk factors from the 2020 report moved to mid-life (physical activity, diabetes, depression, and smoking), and an additional two risk factors have been included, one of which (high LDL/cholesterol) is in the mid-life risk factors, and the second (visual loss) is a later life risk factor [97].

There has been a surge in dementia risk modelling [244, 245, 198, 182, 246, 97, 247, 248, 249, 250, 251, 252, 253]. An updated systematic review in 2024 included $n=74$ new publications since the previous systematic review in 2015 (fourteen of these studies used electronic health records in some format) investigating dementia risk prediction models [253]. This systematic review [253], and also other authors [254], caution against the use of these in the general public. The concerns

related to dementia risk models' validity, and how they translatable to different population groups. One opinion piece highlighted the need to still understand the "*the disease before dementia*" period as important before dementia risk prediction models can be fully understood [254].

One such model specifically for mid-life risk factors in dementia is the Dunedin Alzheimer's Disease Related Dementias Risk Benchmark (DunedinARB) [224]. Researchers collated mid-life risk factors from a number of studies: The cardiovascular Risk Factors, Ageing and Incidence of Dementia (CAIDE) [246], The Lifestyle for BRAinhealth (LIBRA) [244], The Australian National University Alzheimer's Disease Risk Index (ANU-ADRI) [255], and the Lancet Commission 2020 (Lancet) [182] to create a risk index for Alzheimer's disease related dementias (ADRD).

The DunedinARB included ten domains encompassing different risk factors: 1) Genetic risk (*APOE* status), 2) Lifestyle risk (Physical activity, alcohol consumption, tobacco smoking, diet), 3) Socioeconomic status risk (Education), 4) Psychosomatic risk (Depression, cognitive engagement, social engagement/isolation), 6) Cardio-metabolic risk (BMI/obesity, hypertension, high cholesterol, diabetes, coronary heart disease, chronic kidney disease), 7) Inflammatory risk, 8) DNA methylation ageing risk, 9) Harmful events and exposures (Traumatic brain injury, air pollution exposure, pesticide exposure), 10) Subjective overall health risk [224]. Additionally, risk factors of age and sex were considered. The recent Lancet Commission on Dementia Prevention 2024 [97] suggests further modifications in potentially modifiable mid-life risk factors since the DunedinARB [224] was produced. This highlights one complexity in developing dementia risk prediction modelling as new research information on risk factors is updated.

In summary there are many different models of dementia risk and I have highlighted one specific to mid-life as this is a period in a person's life where possible pre-clinical signs may be developing. Moving towards an index of dementia risks

in models can help with public health, at a population and individual level, to modify these risk factors. As mentioned, there are concerns about validity, generalisability and translatability both within an individual country population and in other countries with different ethnicities [253]. Identifying risk factors at critical or sensitive periods of a person's life-course in terms of dementia risk is also important. Understanding the causal pathway, not solely association, is complex for many of the risk factors in dementia, and may vary due to where an individual lives.

1.5 Location, ageing and cognition

In this section, I discuss location and how this may be important for our health in general as well as for our overall ageing and our cognition. I start by defining the terms 'health' and 'medical' geography, then discuss location with a focus on geography and brain health, highlighting key considerations in models that encompass ageing, geography and cognition, differences in risks between rural and urban living and interaction between the 'gene-environment.' In [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#), I additionally discuss rural or urban living with a specific focus on influencing dementia risk.

1.5.1 Defining the terms health and medical geography

Health and medical geography are two terms that can, for some, be interchangeable. Medical geography considers the geography of health and disease in a quantitative, rather than a specific 'medical' approach [256, 257, 258, 259, 260]. It does this by considering the spatial patterns of disease, disease mapping, disease ecology, and possibilities for why disease occurs at specific places and times [258, 256, 257, 260]. In contrast, health geography typically refers to three central constructs: place, health and well-being [257, 256, 261, 262]. In the context of dementia, the construct of well-being is important both in terms risk factors for dementia, and also a key component of living with dementia. There were changes

in the 1990s, with the addition of social theories to the concept of medical geography that slightly blurs the distinction between medical and health geography [258, 256, 257].

Broadly speaking, both can be compared as the geometric constructions of space in medical geography, versus constructed and experimental aspects of place in health geography. Both terms are commonly often used interchangeably and are dependent on the background of the researcher. Traditionally, medical researchers used quantitative methods such as statistics and epidemiology in conjunction with disease mapping techniques to inform their discipline. Typically these studies adopted a very positivist framework and used the location information to confirm associations.

With the growing interest in the ‘place’ or ‘location’ we live in influencing our health and well-being and the interplay of the individual, and their environment being a dynamic two-way process. There is increasing medical research considering the holistic geographical approach of health and medical geography.

1.5.2 Location: geography and brain health

In this section I discuss models that consider different epidemiological study designs that take a geographical approach and also how the environment/geography can influence brain health, ageing, cognition and dementia risk in general and possible differences between rural and urban areas that might impact on dementia risk. In this context when I refer to ‘brain health’, it is as described by the WHO in Section 1.4.3 (Brain health). It refers to the state of brain functioning for an individual to allow them to reach their potential irrespective of disease status. This is different to how location may influence dementia risk. However, where you live may influence an individual’s ability (or opportunity) to optimise their brain functioning. Geography and brain health are separate but co-related issues.

Geography can refer to the geometric space in medical geography or the concepts

of place, health and well-being in the geography of health. All these need to be considered when investigating geographical determinants in relation to brain health, ageing and dementia risk and environmental factors that might be influencing that risk. A recent narrative review highlighted a number of environmental factors linked with Alzheimer dementia risk [63]. These included potential risk factors such as: metals/metalloids (particularly aluminium, arsenic and lead), pesticides, vitamin D deficiency, western diet, electromagnetic fields, infections and inflammation, gut microbiota dysbiosis, depression, social isolation, low educational levels, hearing impairment, physical inactivity, smoking, obesity, hypertension, diabetes, alcohol abuse and air pollution [63].

Additionally, potential protective factors were listed such as Antioxidants, Mediterranean diet, fruit/vegetables, caffeine, physical activity, anti-inflammatory drugs, long duration of education, cognitively demanding work [63]. Several of these were also highlighted by the 2024 Lancet Commission on Dementia Prevention: depression, social isolation, low educational levels, hearing impairment, physical inactivity, smoking, obesity, hypertension, diabetes, alcohol abuse and air pollution [63, 97].

In comparison, a previous systematic review of environmental risk factors and increased risk of dementia found evidence on meta-analysis for only these environmental risk factors: air pollution, aluminium, silicon, selenium, pesticides, vitamin D deficiency; and electric and magnetic fields [263]. This systematic review did not consider health risk factors (i.e. depression, obesity, activity levels) or social risk factors (i.e. social isolation, low educational levels) as environmental risk factors [263]. This highlights an important point that it is not universally accepted what constitutes an ‘environmental’ risk factor.

In the context of this thesis I will include all (traditional geographical/environmental as well as health and social) as environmental risk factors, as I consider differences between rural and urban areas that might be influencing health and risk of

dementia. This is because an individual, and their environment (location), are intricately linked.

To investigate the interplay of disease frequency and geographical location, there are different epidemiological study designs. Epidemiological studies investigating a group, rather than individuals, are known as ecological studies, and typically studies investigate a potentially explanatory variable and a disease frequency within and between locations [256, 191, 193]. Typically, in ecological studies the data available would be whole population data (e.g. alcohol consumption) but not at an individual level to do detailed analysis.

Migrant studies are naturally occurring experiments as they investigate the changes in a disease and risks between groups of people migrating from one area to another [256, 191, 193]. This can highlight important aetiological factors at different life course time points, but are hugely complex to undertake and investigate this. In relation to other neurological diseases, one historical example is understanding the risk for Multiple Sclerosis (MS) on individuals migrating to South Africa from Northern Europe, and how the future risk of MS was reduced if they emigrated before the age of fifteen [264]. However, this has been disputed in other countries not suggesting an increased risk before the age of fifteen, which highlights some of the difficulties in migrant studies [265].

There are limitations with epidemiological studies investigating geographical variations of diseases and how the geographical location of where we live influences our health and disease. For ecological studies, a limitation is the ‘ecological fallacy’, in which results or observations at a group level may not apply at the individual level [256, 191, 193]. Migration is often considered to be biased by self-selection for various reasons and therefore may not be translatable to all individuals.

The environment is one aspect of geography. Another perspective to consider in the interaction between location, health, ageing and cognition is an ‘ecological

model’ such as the ‘press-competence model’ first proposed in 1970s by Lawton and Nahemow [266]. The model refers to the balance of environmental pressures with individual competence and how this influences a person’s ability to adapt [266]. It is a balance of exposure to environmental pressures combined with adequate support and stimulation and individual competencies to reach an ‘adaption level’ [266]. Lawton’s model further emphasises the complex interplay of the environment and a person over time in relation to successful ageing [266].

This model is founded on the consideration that: *"as people age, their competence is reduced due to losses in functionality and therefore they are more subject to environmental demands, a condition defined as ‘environmental docility’"* [267]. Across an individual’s life-course, it has been proposed that old age is a critical time period where individuals are most affected by their physical environment [266]. A prior published review discussed the environment and cognition in the context of this ecological model, and proposed that *"human behaviour results from a dynamic interplay between individuals and their social and physical environments"* [268].

This review also discussed the environment in relation to ageing [268]. It highlighted that *"with ageing, the experience we have of the environment is reshaped both by physical, sensory, and cognitive changes, and by modifications of the perceived affordances offered by the environment"* [268]. There is a complex interplay between the environment a person lives in and their health [269, 268].

The term ‘environment’ does not focus solely on the geographical area but the type of environment in relation to the physical, structural, socio-economic, educational, supportive factors within that area, along with many other influencing factors, some which might yet be unknown. The interaction between a person and the place they live in is multi-modal, and is particularly important in relation to cognition and brain health [268]. This can be illustrated clearly by considering the direct and indirect associations between the environment and cognition [268].

Figure 1.8 (reproduced from Cassarino et al. 2015 [268]), outlines a flowchart of variables to consider which may be underpinning the environment-human interaction, and subsequent impact on cognitive ageing.

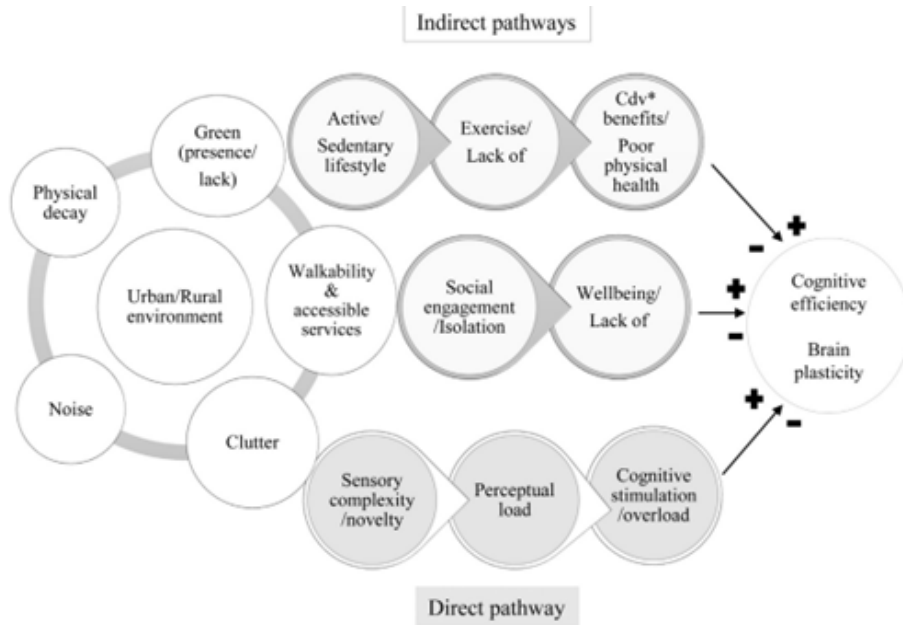


Figure 1.8: Direct and indirect associations: the environment and cognition [268]

Source: Reprinted from *Ageing Research Reviews*, Vol 23, Authors: Marica Cassarino and Annalisa Setti, Title: *Environment as ‘Brain Training’: A review of geographical and physical environmental influences on cognitive ageing*, Pages No. 169, Copyright (2015), with permission from Elsevier and Copyright Clearance Center.

In conceptualising how environmental drivers might be involved as geographical determinants of ageing and brain health, it is helpful to consider what many believe are shared common underlying cellular and molecular mechanisms of ageing in age-associated diseases. Figure 1.9, reproduced from Pandics et al. 2023, illustrates a conceptual model of environmental drivers towards unhealthy ageing and the underlying mechanisms [270]. Some mechanisms include: oxidative stress, cellular mitochondrial and energetic dysfunction, impaired cellular stress resilience, genetic instability and DNA damage, cell senescence, inflammation, epigenetic changes and many others as illustrated in the following Figure 1.9 [270].

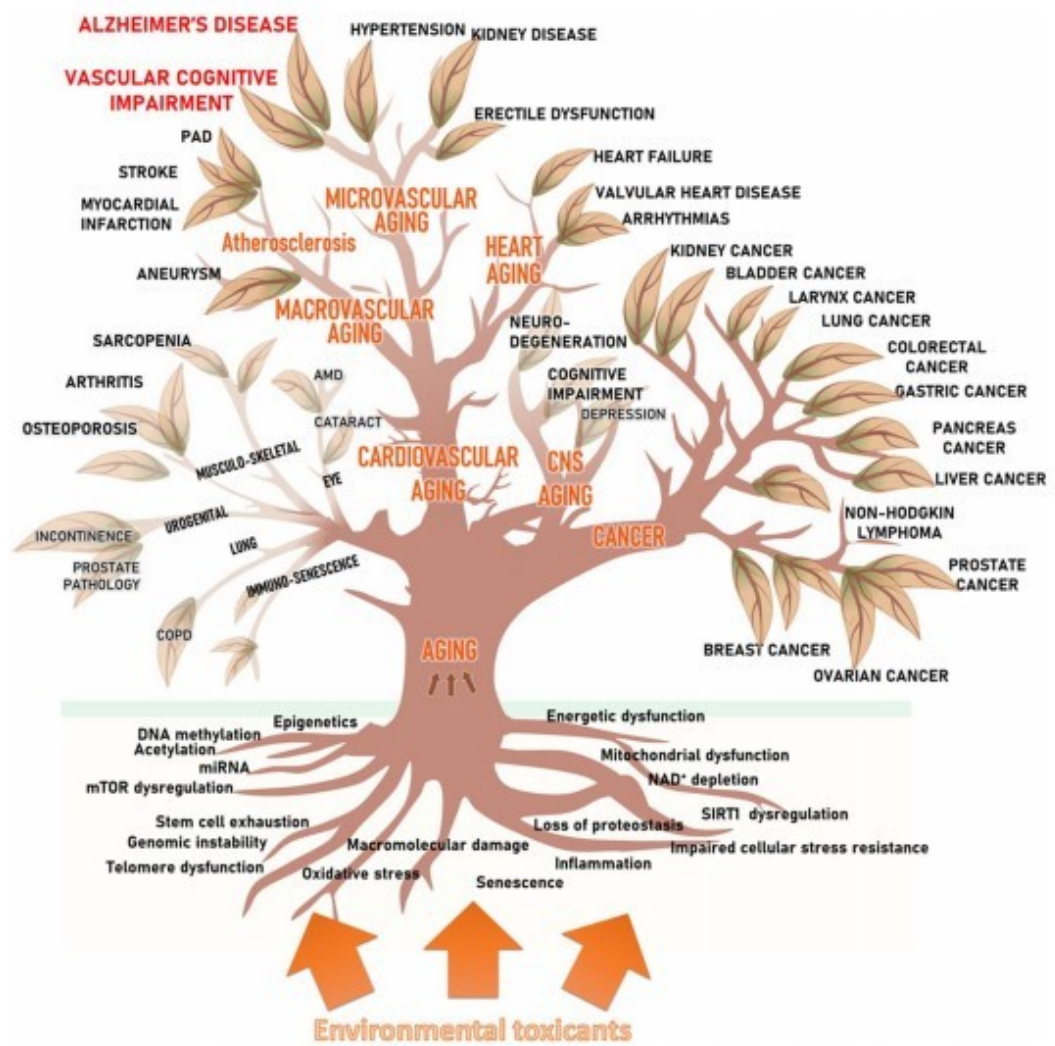


Figure 1.9: Conceptual model of environmental drivers and unhealthy ageing [270]

Source: Reproduced from GeroScience, Authors: Tamas Pandics, David Major, Vince Fazekas-Pongor et al., Title: *Exposome and unhealthy aging: environmental drivers from air pollution to occupational exposures*, 2023. Used with permission under the Creative Commons Attribution 4.0 licence for this article. The Figure without accompanying title and footnote is included from the original article.

In 2005, the term ‘*Exposome concept*’ was introduced [271]. This concept encompasses environmental exposures and accompanying biological responses throughout the life-course that impacts on health, well-being and risk of disease [272, 271,

273, 274]. It refers to every exposure that an individual encounters from conception to death, and relates to the ‘non-genetic’ exposures [271, 273, 272, 274, 275, 276]. It is an approach to address exposure-health association. Although almost twenty years has passed since this concept was suggested, there are difficulties in tools to study the health impacts of complex environmental exposures across an individual’s life [272, 273, 271, 274, 276, 275]. There is additional complexity, because although it encompasses the ‘non-genetic’ exposures, the role of epigenetics and environmental exposures suggest there is a complex inter-web between life-course environmental exposures and genetics.

World urbanisation and health

The world is rapidly urbanising [277]. Urbanisation is defined as: *“the increase in the proportion of people living in towns and cities. Urbanisation occurs because people move from rural areas (countryside) to urban areas (towns and cities). This usually occurs when a country is still developing [278, 279].”*

The United Nations state that approximately half of the world’s population live in urban areas, and that this is expected to rise to 70% by 2050 [277]. The world’s urbanisation brings many changes to our natural environment. The location we live in is changing both in structure (spatially) and also over time (temporally). With urbanisation, access to healthcare services, education and facilities is changing but not equitably [277]. Location influences our health and disease risk, not solely due to the accessibility to healthcare services, but through a far more complex inter-web of variables contributing risk or protective factors that are not yet fully understood. It is impossible to identify all the geographical (or location) determinants potentially contributing to brain health.

In contrast, there is a term called ‘counter-urbanisation’ where people move from urban areas to more rural areas which can be seen more in HICs (e.g., individuals choosing to retire to a more rural setting, accessibility to rural areas improving, or opportunities to work remotely) [278].

Traditionally the distinction between a rural and urban location encompasses a location's population size, density and accessibility (or accessibility to the nearest urban area). However, defining what constitutes a rural area is complex. Even within the UK, the definition of rural and urban locations (and the parameters used to define this) varies across the four nations. This will be discussed and explored further, including additional differences in the definition of rural, in Chapters [2](#), [6](#) and [7](#).

Considering differences between rural and urban areas that might be driving risk of developing dementia, then some risk factors may vary between the two areas. Geography encompasses more than just the spatial environment. It encompasses the social environment an individual lives in, and the environmental risk factors. For dementia, this could include traditional environmental risk factors such as air pollution, vitamin D levels due to latitude, environmental toxins [[263](#)], or it could also include social determinants such as social interaction, accessibility to services, access to green space [[280](#)].

Many of these risk factors may be altered (or influenced) by living in a rural or urban location. It may not always be in the direction expected. For instance, living in an urban area which is highly populated may imply there is more potential for social interaction. However, in a rural area the population may be less but there may be a more meaningful network of social interaction. Conversely, the urban environment might create physical barriers in the built environment, particularly those considered to be in 'vulnerable' groups, which is why a recent focus in urban settings includes looking at reducing *"social isolation and build social connectedness through human-centred, inclusive urban planning"* [[281](#)]. Furthermore, the population demographics may traditionally differ between urban and rural settings creating differences between individuals.

Traditionally, urban areas may attract people for jobs and opportunities, being

areas where universities and employment are largely found. A document [282] from University College London discussing the ‘*The Ten and a Half Myths that may Distort the urban Policies of Governments and International Agencies*’ highlights what it describes as a ‘semi-myth’ that “*most poverty is in rural areas*”. This states that whilst this is: “*not a myth globally since more poverty is still in rural areas*”, importantly the authors draw distinctions from different areas in the world suggesting that this statement is no longer the case in Latin America, Europe or North America, but remains the case for most African and Asian countries [282].

When considering differences between rural and urban areas, and the impact on an individual’s health, there is focus on the ‘natural environment.’ Indeed, it could be argued that the natural environment is a pre-requisite for health, and a fundamental contributor to liveability. Liveability is beyond solely the physical setting and includes social interactions and the social determinants of health [283]. The natural environment typically refers to areas which are not man-made, or the environment without human activity, such as green space and natural habitats [284, 285].

Green space exposure is an area of growing interest in relation to an individual’s health and also specifically for dementia risk [286]. Increased green space exposure (exposure to natural environments) has been proposed as a potentially modifiable factor in urban environments for subsequent risk of dementia through influencing a number of factors (neurological health, psychological well-being, social interactions, stress, physical activity and sleep quality) [286]. This study found the Hazard Ratio (HR) for dementia was 0.97 (95% CI: 0.96-0.98) for each Interquartile range (IQR= 0.12) increase in green space, demonstrating that increased green space reduced incidence of dementia [286]. Additionally, this study also showed protective associations between increasing green space and all-cause mortality $HR = 0.93$ (0.92-0.93) [286].

Many studies investigating the natural environment and health research in terms of green space use the Normalized Difference Vegetation Index (NDVI) which is calculated from satellite imagery which senses the quantity of green vegetation and assigns it a score [286, 287]. There are difficulties with this method in interpreting different types of canopy cover in urban settings and what the NDVI values mean in these settings [287]. Nonetheless, there is a growing interest in green space in relation to variation in rural and urban living and impact on health and risk of disease. By default, rural living is more likely to have increased exposure to green space than urban living.

Rurality/Urbanicity and dementia risk

There are differences in rural and urban area for multiple health conditions. How this impacts on future dementia risk is the topic of this thesis. There is evidence for risk factor differences between rural and urban areas for other psychiatric diseases, or for an individual's physical or psychological well-being. Some of these are captured in dementia risk prediction modelling [253] as mentioned in Section 1.4.4, and include three commonly used variables for dementia risk (depression, obesity, physical activity).

The role of an urban versus rural environment on a person's psychological well-being, including depression, has been investigated in more detail. Several papers report that the urban environment is linked with increased risk of depression [288, 289, 290, 291, 292, 293]. This may also be in part be due to deprivation or social isolation. Other published papers have considered the neurological/neurobiological impact of urban living compared with rural living [294, 295]. One paper used the term 'neurourbanism' specifically in relation to urban living impacting on our mental health [296]. The authors defined the term as it *"looks to understand the relationship between urban environments and mental well-being and is well placed to assess the role of these environments on the urbanised and ageing global population"* [296].

Several studies from China suggest the reverse. These studies report that depression is lower in urban areas [297, 298, 299]. The authors attribute this to individuals migrating from rural to urban China due to socio-economic advantage. The studies conclude that having an urban household, and possibly access to private pension, are an independent protective role in shaping mental well-being [297, 298, 299].

Whilst rural or urban living may influence depression, and overall mental health, this may vary between HICs and L-MICs. The effect of the impact on greenspaces in L-MICs was recently highlighted in a scoping review which demonstrated a sparsity of publications in LMICs investigating this [300]. Another risk factor for dementia is high BMI/obesity in terms of increasing cardiovascular/cerebrovascular causes of dementia [97]. This risk could potentially vary between rural and urban environments, potentially influenced by physical activity and other factors. There may or may not be overlap with this risk factor and level of physical activity with a rural environment creating a more ‘active’ lifestyle.

One study suggested that a 90-minute walk in nature (grassland and trees) versus urban environment was associated with reduced levels of rumination (which has been linked with depression), so there may be links between several of these risk factors for dementia and differences between rural and urban settings [301]. Another study investigated the impact of walking in different urban environments (urban busy/urban green/urban quiet) on brain activity in older people focusing on changes in alpha (associated with relaxation), and beta (associated with attention) brain activity recording using mobile electroencephalography (EEG) [296]. It found that the greatest difference was between the urban busy and urban green sections [296].

In a 2024 systematic review of 74 studies investigating dementia risk prediction models [253], only one study used rurality in their prediction model [302]. This was in Mexico, a country classed in 2023 as Upper Middle Income country ac-

cording to The World Bank classification [303]. The relevance to countries of different incomes is unknown.

A recent study investigated the social determinants of health amongst older adults already living with dementia in urban and rural areas [280]. This study found those living in urban areas were more likely to have a higher level of education, a known risk factor for dementia (college degree $p < 0.001$), than those living in rural areas [280]. Conversely, those living in rural areas were at most risk of financial hardship ($p = 0.003$) [280]. The authors found no significant differences in other domains of social determinants of health such as physical inactivity, social isolation or stress between rural and urban areas [280]. Importantly, this was amongst adults already living with dementia rather than investigating risk factors for developing dementia that differ between rural and urban settings [280].

Cognitive functioning and the environment

Cognitive functioning is influenced by our brain health and cognitive maintenance. It is also influenced by cognitive reserve. Cognitive reserve is the brain's ability to adapt, or cope, with changes. Cognitive reserve is important for future risk of dementia [304, 305, 306, 307, 308]. People with higher cognitive reserve can have reduced dementia risk. A large component of cognitive reserve, and future dementia risk, is due to early life factors such as education (and higher level of education) [309], and shaped by prenatal and early postnatal factors [310]. Cognitive reserve can be influenced by genetics, male/female sex, but also importantly influenced by the environment, and geographic determinants on brain health [306, 311, 295, 312].

Depending on the nature of the location we live in, the environment/geography might provide facilities to enhance cognitive functioning (performance), or cause cognitive overload depending on both the levels of stimulation and the individual's cognitive and physical abilities [313, 314, 315]. This in turn might influence the individual's cognitive reserve to then adapt to further changes. The Attention

Restoration Theory (ART) is a theory whereby individuals can concentrate better after spending some time in nature or being exposed to scenes of nature thereby having a positive impact on cognition, in part by providing rest from attentional fatigue [316, 317].

For an environment to be restorative it must have some level of effortless/‘soft’ fascination - one example would be natural green spaces - and a reduction in ‘hard fascination’ which can create demands on an individual, commonly found in urban environments [314, 296, 318, 317, 319, 320, 321]. Conversely, another study investigating ART in older adults demonstrated no attention restoration effect changes in a population of older adults with before and after exposure to natural or urban environment scenes [322].

It remains unclear at what time point this has most impact on an individual in their life-course. It is helpful to consider the ART in relation to overall active ageing and brain health, as our brain changes as we age, and can be influenced by our activity (physical and cognitive). Active ageing is a term described by WHO as: *"Active ageing is the process of optimizing opportunities for health, participation and security in order to enhance quality of life as people age"* [323].

Gene-environment interactions

Whilst there are individual environmental risk factors that might influence the risk of dementia, and which could vary between rural and urban areas, there is another component of the individual’s interaction with their environment of importance: the gene-environment theory [63]. Our genome might not change, but our epigenome may be influenced, or ‘modified’, by our environment, thereby environmental risk factors for dementia affecting an individual through this route [324].

A past review has highlighted that research has previously shown higher risk of mental disorders among people living in urban than rural areas [325]. The authors

of that review also proposed that epigenetic changes (via DNA methylation) may be a potential mechanism by which *"features of the urban environment contribute to psychopathology"* [325]. They propose that environmental features that could be more prevalent in urban environments (environmental toxins, including second hand smoke and air pollution) and are known to affect cognitive function, particularly in childhood, so may be involved in this difference [325]. The review proposes that as lower cognitive function in childhood can be linked with predicting risk and severity of some mental health disorders in adulthood then the urban environment might impact on risk for mental disorders via impacting on the developing brain [325].

In view of this ‘gene-environment’ interaction, some environmental risk factors for dementia have been considered in this context [63]. For instance, diet, gut microbiome, lifestyle, alcohol, smoking and air pollution have been proposed to have gene-environment interactions in the risk of dementia in a recent review [63]. Additionally, some of these (e.g. gut microbiome and air pollution) have been considered to vary between rural and urban areas, and some authors propose a potential causal pathway via epigenetics [326, 327].

A number of studies have shown a positive association between air pollution and risk of neurodegeneration [182, 328, 329, 330, 331, 332, 333, 334]. A systematic review (2019) examined a number of pollutants associated with air pollution and found that greater exposure to particulate matter, $PM_{2.5}$, nitrogen dioxide, nitrous oxides and carbon monoxide were all associated with increased risk of dementia [328]. Other studies have shown differences between men and women and this risk suggesting that women living in places with higher levels of fine particulate matter ($PM_{2.5}$) had increased risk of cognitive impairment and dementia than men [335].

In relation to rural versus urban living, then urban areas would traditionally be linked with higher levels of air pollution, but it is not always as clearly defined

with other potential contributing factors such as the socio-economic status of the area also. A 2017 study in England found that *"relationships between pollution level and socioeconomic deprivation were non-linear and varied by urban-rural status"* [336]. Additionally, this study found that in rural areas, *"the absolute burden of air pollution-related LYs [Life-years] lost was lowest in the most deprived deciles"* [336]. How this translates to other countries, and in particular comparisons between HICs and LMICs, is unclear.

When investigating air pollution in urban environments, another study showed that human and mice brains exposed to particulate urban air pollution showed a global decrease in repressive epigenetic marks, together with increased levels of Alzheimer disease dementia related tau phosphorylation [327]. Another paper suggested a possible mechanism for this, although the study was not focused on differentiating between rural and urban areas, it proposed that exposure to air pollution, particularly wood smoke, as a hypothesis to induce changes in DNA methylation of genes linked to inflammatory pathways central driving the onset and progression of Alzheimer disease dementia [337].

Location across an individual's life-course

Where a person lives across their life time might change, which may be influenced by movement of migration (adding other confounders if representing a self-selected group migrating due to occupation or due to inequalities such as war or famine). Other societal changes may influence movement between rural and urban areas across an individual's life. For example, recently in Scotland, the proportion of the population living in rural areas has increased [338]. This may in part be associated with a 'post COVID-19' era in which working from home or remotely is far more acceptable. Similarly, in large capital cities such as London, where house prices are expensive, there has been movement of young families living further from the city as commuting opportunities improve and to better manage increased cost of living [339].

Considering a life-course approach, as in Section [1.4.1 \(Life-course epidemiology in relation to dementia\)](#), a person’s location may vary at different time points which in turn may influence an individual’s risks of developing dementia across their life-course. Evidence suggests that cognitive abilities in both childhood and adulthood are important predictors of life outcomes: health, morbidity and mortality [[220](#), [217](#), [213](#)]. Childhood is often considered, in life-course epidemiology, as a ‘critical period’ in terms of neurodevelopment [[192](#), [193](#), [191](#), [201](#), [200](#)].

There are limited studies investigating location in childhood and adulthood in terms of a rural or urban setting, and dementia risk [[236](#), [235](#)]. Therefore, considering geographical determinants of brain health, and risk of dementia across a person’s life, and over time, is complex as: a) the individual may move between locations across their lifetime, and b) the location is a changeable variable dependent on what constitutes that environment over time and within the context of the historical era being investigated.

1.6 Chapter conclusions

In summary, in this chapter I have discussed an overview of dementia highlighting some of the complex challenges in the definition of dementia and diagnostic approaches. I then discussed dementia prevalence worldwide and why this might be still an under-representation of the true dementia prevalence. Finally, I discussed epidemiology in relation to dementia, including life-course factors, geographical determinants of brain health, and the complex interplay between an individual and their environment with a focus on differences in rural or urban locations that might influence this. I concluded this chapter by outlining the aims and objectives of my thesis. The next chapter [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#) will now systematically review publications investigating geographical variation of dementia and consider whether there is a rural/urban variation in dementia prevalence.

1.7 Thesis research question, aims and objectives

My research question for this thesis is *‘Does dementia vary between rural and urban areas and how can we use routinely collected health data to answer this question?’*

The aim of this thesis was to consider whether a rural or urban location influences risk of dementia by using routinely collected health data.

The objectives of this thesis were threefold:

1. To systematically review the literature of studies investigating geographical variation of dementia between rural and urban areas. To undertake meta-analysis, where possible, on the studies and establish if there are different odds ratios reported between rural and urban areas for dementia prevalence.
2. To undertake novel research using nationally collected routine data in Aotearoa New Zealand (ANZ) within the Integrated Data Infrastructure (IDI) and investigate the dementia prevalence in ANZ between 2012-2015 and the geographical variation.
3. To undertake novel research using a historical population-wide survey (the Scottish Mental Survey 1947) linked to routinely collected data and investigate who developed dementia and whether living in a rural or urban area influenced the development of dementia.

1.8 Overview of PhD

This thesis begins with a systematic review and meta-analysis of geographical variation of dementia in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#), focusing on rural versus urban areas. In [Chapter 3: Using big data for dementia research](#), I consider the role of big data and routinely collected healthcare data and its secondary uses for research. In [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using rou-](#)

tinely collected data, I produce novel research by using routinely collected big data in Aotearoa New Zealand and investigate this at a population level to consider where dementia counts are spread geographically between North and South Island. In [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), I discuss the methodology for my research in Chapter 6 to investigate geographical variation within the historical population-wide Scottish Mental Survey 1947, SMS1947, linked to routinely collected health data. In [Chapter 6: Results of the SMS1947 dementia data linkage study](#), I produce novel research investigating whether rural or urban living influences dementia diagnosis in the SMS1947 linked to routinely collected health data.

Chapter 2:

Systematic review and meta-analysis of rural or urban living with risk of dementia

2.1 Chapter introduction

The first step in investigating my thesis research question (*‘Does dementia vary between rural and urban areas and how can we use routinely collected health data to answer this question?’*) was to understand the current opinion in literature by undertaking a systematic review, with meta-analysis where sufficient data were available. This chapter is focused on this systematic review which has been uploaded to *medRxiv* [1] and is in the process of being updated for submission to a peer-reviewed journal. I now give a brief introduction followed by the manuscript, and then discuss some key areas of the systematic review in more detail in the subsequent sub-sections of this Chapter.

As discussed in [Chapter 1: Introduction](#), understanding the environment or geographical location a person lives in across their lifetime and their subsequent risk of dementia is important. It may highlight possible risk or indeed protective, or modifiable, factors for dementia. Non-random geographical variation in dementia risk could also have major implications for the distribution of dementia-related health, social and support services as well as consideration of modifications to the environment that we live in to reduce the risk of dementia.

A past systematic review investigated many different geographical scales in relation to dementia prevalence and found an increased risk of dementia in rural

than urban areas [340]. This chapter aims to further explore this by undertaking a systematic review of the current literature investigating rural and urban variation of dementia.

2.2 Research aims and objectives

The objectives of this systematic review were twofold:

- To systematically review the literature for observational studies that investigated dementia incidence or prevalence and rural/urban location.
- To extract odds ratios and undertake meta-analysis, where possible, to investigate rural versus urban living and risk of dementia

2.3 Published manuscript on *medRxiv*

This chapter includes the manuscript that has been uploaded on medRxiv [1], which is in the process of being updated for submission to a peer-reviewed journal:

- **Walesby, KE**, Shenkin SD, Burton JK, Dunlop C, Fenton C, Gray D, McGillicuddy LA, Starr JM, Wilkinson T., Muniz Terrera G., Russ, TC. Geographical variation in dementia: systematic review with meta-analysis. *medRxiv* 2023.12.11.23299178; doi: 10.1101/2023.12.11.23299178 [1].

I include my registered protocol for my systematic review on PROSPERO (CRD42016050323) which was updated to reflect the focus on rural versus urban variation in [Appendix 1](#) [341, 342]. Initially, the aim was to investigate two levels of geographical locations in the analysis: 1) country to country variation, and 2) rural and urban variation. However, there has been a substantial increase in the publications in dementia research and geographical variation since the past systematic review (search undertaken in 2010) and this systematic review search

on 31st October 2018. Therefore, it was more appropriate to focus on rural and urban comparisons as this was the research question in my thesis.

I include the *medRxiv* pdf copy of the manuscript [\[1\]](#) followed by the Appendices 1-7 from the manuscript in the body of this Chapter for clarity.

Following the manuscript and appendices, I discuss how this research contributes to existing literature, key themes from my systematic review, additional discussion since my systematic review, future work and concluding remarks.

Geographical variation in dementia: systematic review with meta-analysis

Walesby K.E.^{1*}, Shenkin, S.D.², Burton J.K.³, Dunlop C.⁴, Fenton C.⁵,

Gray D.⁶, McGillicuddy L.A.⁷, Starr J.M.¹, Wilkinson T.⁸, Muniz Terrera G.⁹, Russ T.C.¹

¹Alzheimer Scotland Dementia Research Centre and Centre for Cognitive Ageing and Cognitive Epidemiology, 7 George Square, University of Edinburgh, UK

² Ageing and Health Research Group, Centre for Population Health Sciences, and Advanced Care Research Centre, Usher Institute, University of Edinburgh

³ Academic Geriatric Medicine, School of Cardiovascular and Metabolic Health, College of Medical, Veterinary & Life Sciences, University of Glasgow

⁴ Old Age Psychiatry, Edinburgh Royal Infirmary, NHS Lothian, UK

⁵ NESSIE, Usher Institute, University of Edinburgh

⁶ Fairnington Centre, Hexham General Hospital, Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust (CNTW), UK

⁷ Old Age Psychiatry, Royal Edinburgh Hospital, NHS Lothian, UK

⁸ Centre for Clinical Brain Sciences, University of Edinburgh, UK

⁹ Department of Social Medicine, Ohio University, Ohio University Heritage College of Osteopathic Medicine and Edinburgh Dementia Prevention, University of Edinburgh, UK

*Corresponding author. Alzheimer Scotland Dementia Research Centre, 7 George Square, University of Edinburgh, EH8 9JZ. E-mail:

ABSTRACT

Background: Understanding geographical variation of dementia could highlight important modifiable socio-environmental risk factors. A previous systematic review (2012) identified an increased risk of Alzheimer dementia in rural living in High-Income Countries (HICs), with a dearth of studies in Low to Middle-Income Countries (L-MICs). We updated this review to examine geographical variations in dementia, to encompass the growing number of studies in this field.

Methods: We systematically reviewed the literature for cross-sectional or longitudinal observational studies that compared dementia incidence or prevalence between two or more geographical areas including rural and urban settings.

We conducted a narrative synthesis of included papers. Where possible, we undertook meta-analysis, generating odds ratios for rural versus urban dementia prevalence and stratified the analysis by HICs and L-MICs.

Results: We identified 38 relevant papers, encompassing approximately 98,502,147 people. Twenty-seven papers were included in the quantitative synthesis. Study methodologies varied widely. Dementia rates varied geographically (0.43-38.5%). Overall, rural living was associated with small increased odds of dementia (OR, 1.20, 95% CI 1.03-1.40; P value = 0.0182). Stratification by HICs and L-MICs demonstrated further variation, with increased odds of dementia in rural areas in L-MICs but not HICs.

Conclusions

There is some evidence of geographical variation of dementia. Rural living was associated with small increased odds of dementia, with stratification showing evidence in rural areas of L-MICs but not HICs. We believe this has not been reported previously. Future research

must consider life course geographical exposure and addressing heterogeneity in definitions of ‘rural’ and ‘urban.’

Keywords: Dementia, Alzheimer disease, epidemiology, geography

What this study adds

We confirm that rural living (compared to urban living) is associated with a small increased odds of dementia (OR 1.20, 95%CI 1.03-1.40). We demonstrate for the first time that this is driven by increased odds of dementia in rural areas in Low to Middle-Income Countries (LMICs) rather than High Income Countries (HICs), and that the odds of dementia were higher in urban areas in large studies in HICs. Future studies need to carefully consider study setting, method of dementia ascertainment, when exposures may occur, and risk of bias, to understand the role of environment and geography in dementia risk.

INTRODUCTION

Dementia is a growing worldwide public health concern. The epidemiology of this complex syndrome is not yet fully understood.¹ Age and known genetic, health, and lifestyle risk factors play an important role in dementia incidence, but together they do not fully explain dementia aetiology. Life course geographical variation might influence dementia risk or protective factors.

A systematic review in 2012 found “evidence of geographical variation in rates of dementia in affluent countries at a variety of geographical scales.”² Furthermore, an increased risk of Alzheimer’s dementia was associated with rural living (Prevalence OR 1.50, 90% CI 1.33-1.69), with early life rural living further increasing this risk to approximately double (OR 2.22, 90%CI 1.19-4.16). However, these results were based on very few studies, and there

have been several subsequent studies. We therefore updated this systematic review² to reflect the developing body of evidence in this field. In our present systematic review, we report a comprehensive overview of the literature specifically examining the influence of rural and urban living on risk of dementia.

METHODS

Registration and reporting

We prospectively registered the protocol with PROSPERO, reference CRD42016050323 (<http://www.crd.york.ac.uk/>) and report the review in accordance with the Preferred Reporting of Items in Systematic Reviews and Meta-Analyses (PRISMA) guidance.^{3,4}

Information sources

We searched multiple databases on 18th-31st October 2018 for papers since the published review² search date of 15th April 2010: search information and indicative search strategies were developed with an Information Specialist (CF), (Appendix 1).

Eligibility criteria

Eligible studies included: cross-sectional or longitudinal studies of any duration comparing primary dementia rates (prevalence or incidence) between two or more different geographical sites. Our initial search was broad then focused to rural/urban; therefore, many studies (117) were excluded as ‘wrong comparator.’

We excluded studies purely on young onset dementia (<60 years old) or mild cognitive impairment. We did not restrict to English language (Appendix 2).

Study selection and data collection

Two reviewers (from KEW, JKB, and TCR) independently screened all titles and abstracts using Covidence,⁵ and two then reviewed full text with conflicts resolved by discussion.

Google translate was used for data extraction from papers not in English.

Two of KEW, LAM, CD, or DG extracted data using a bespoke and piloted form with discrepancies resolved by discussion and reference to the original paper.

Where articles did not report odds ratios for dementia risk in rural and urban areas,⁶⁻¹⁶ we computed these if the relevant raw data were reported or contacted authors^{9,17-20}; data provided by two authors.^{18,19} We contacted authors regarding full-text publications from abstracts identified, and overlap between studies. When both unadjusted and multivariable-adjusted odds ratios were reported, we report unadjusted odds ratio. When both DSM-IV and 10/66 method were used for dementia ascertainment, we report the 10/66 results²¹⁻²³ (as this has shown greater dementia ascertainment in LICs and LMICs^{24,25}) except when the 10/66 data were incomplete.²⁶

Strategy for data synthesis

Where sufficient data were available, we conducted a random effects meta-analysis of odds ratios. Post-hoc we decided to explore the data by stratifying by Higher Income Countries (HICs) and Low to Middle-Income Countries (L-MICs)²⁷ (Appendix 3), and further by study size. We recognise this classification of country income can change with time (Appendix 3).

Between-study heterogeneity was assessed by the I^2 statistic. Studies which could not be combined quantitatively were reported in a narrative synthesis.

Risk of bias assessment in individual studies

KEW, SDS, and TCR assessed the risk of bias using an adapted Risk of Bias Assessment Tool for Non-Randomized studies (RoBANS), (Appendix 4), with disagreements resolved by discussion.

RESULTS

The narrative synthesis includes 38 new papers (**Table 1**), 27 of which – including 33 cohorts - could be meta-analysed (**Figure 1**).

Geographical locations and sample sizes

Studies were predominantly in the Americas (N=11) and Western Pacific region (especially China, N=11), and also Europe (N=9) South-East Asia (N=4), Africa (N=3), and Eastern Mediterranean region (N=2) (one study covered three areas). Of the meta-analysis papers, 12 papers (13 studies) were based in HICs and 15 papers (20 studies) in L-MICs.

In the 35 papers that included a sample size, this ranged from 90 to 73,000,000 people, median 2170. Of 27 papers in the meta-analysis, study samples ranged from 300 to 21,624,228 people, median 2162.^{6-11,13,15,16,18,19,21,22,28-39} The definition of ‘rural’ and ‘urban’ was heterogeneous. Too few papers included information on education, deprivation, or other socio-economic factors in relation to rural or urban settings and subsequent dementia risk to include in the meta-analysis. Three papers considered childhood and adulthood rural or urban residence and risk of dementia,^{18,40,41} only one was included in the meta-analysis.¹⁸

Dementia diagnosis

There was considerable heterogeneity in dementia ascertainment methodologies (**Appendix 5, Supplementary Table 1**), mostly using surveys/interviews, clinical assessment and ICD codes. Dementia prevalence ranged from 0.43%-38.5% of the study populations in all studies, and 2.4-21% in the meta-analysis studies. Few identified dementia subtype. Most papers compared dementia against non-dementia or ‘normal’, with some including cognitive impairment/MCI in the ‘no dementia’ group, and others excluding those with MCI.

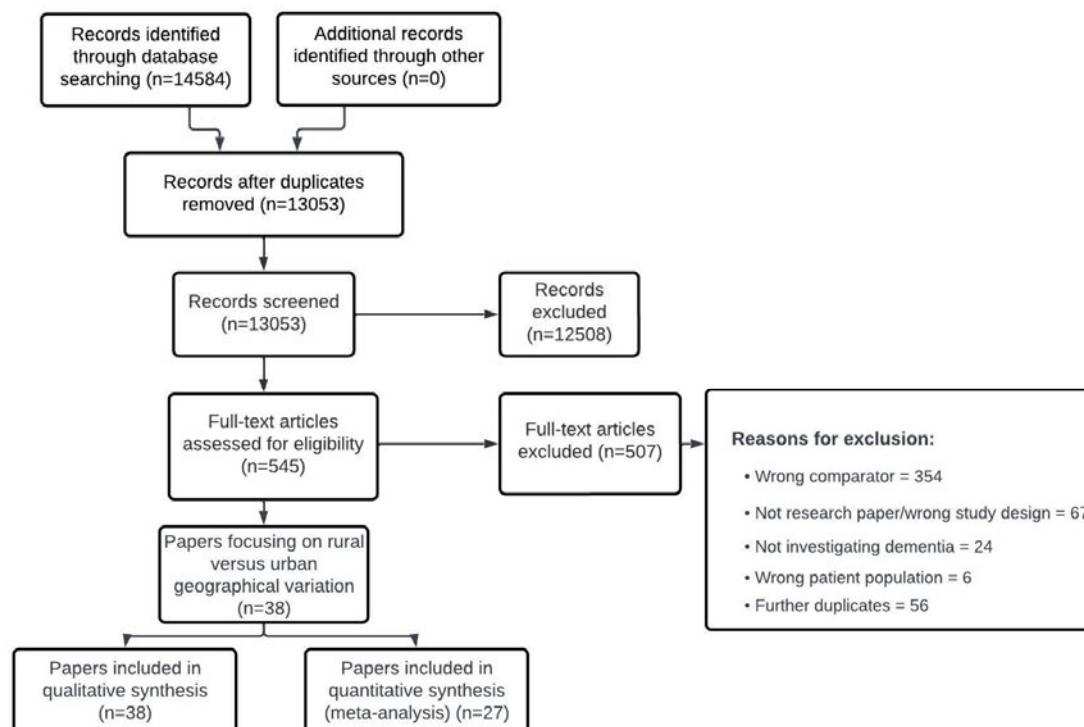


**Figure 1: PRISMA flowchart detailing selection of studies for inclusion in the review:
rural/urban differences in dementia**

Table 1: Executive summary of studies meeting inclusion criteria: rural/urban differences in dementia

Author	Year (data collection)	Country	Age (years)	Study Population		Dementia <i>n</i> (% of total)
				(<i>n</i>)	Urban <i>n</i> (% of total)	
Abner 2016 ^{*28}	2007-2013	USA	Any	1,209,976	766,250 (63.3) Major cities = 139,071 (53.1%)	82,278 (6.8)
Astell_Burt 2018 ^{*29}	2006-2009	Australia	>=45	261,669	Inner regional area = 92,734 (35.4)	3046 (1.2)
Chammartin 2016 ⁴²	2008-2012	Switzerland	Any	60,388	N/A	1,510 (2.5)
Chen 2017 ^{30*}	2001-2013	Canada	55–85	2,066,639	1,674,591 (81.0)	257,816 (12.4)
Chen 2012 ³¹ Anhui study (a)*	2001-2003	China	>=65	2,917	1,736 (59.5)	210 (7.2)
Chen 2012 ³¹ Four provinces study (b)*	2008-2009	China	>=60	3,327	1,787 (53.7)	341 (10.2)
Contador 2015 ⁴¹	1994-1998	Spain	>=65 (mean 72.9 +/- 6.1)	2,711	852 (31.4)	91 (3.4)

Author	Year (data collection)	Country	Age (years)	Study Population (n)	Urban n (% of total)	Dementia n (% of total)
Cornutiu 2010 ⁴³	1980-2006	Romania	Any	Unclear	N/A	N/A
Deng 2018* ³²	Oct-Dec 2015	China	≥60	1,781	464 (26.1)	195 (10.9)
Fronza da Silva 2013 ⁴⁴	July-Oct 2007	Brazil	≥60 (mean 72.3 +/- 7.7)	229	131 (57.2)	11 (4.8)
Ganesh 2010 ¹⁷	2003	USA	any	31,660	N/A	427 (1.3)
Goncalves-Pereira 2017* ²¹	N/A	Portugal	≥65	1,405	690 (49.5)	DSM IV: 51 (7.4)
Goodman 2017* ³³	2011-2013	USA	≥68	21,624,228	16,477,662 (76.2)	3,110,654 (18.9)
Guerchet (a)		Central African				
2013* ²⁶	2011-2012	Republic	≥65	973	500 (51.4)	72 (7.4)
Guerchet (b)		Republic of				
2013* ²⁶	2011-2012	Congo	≥65	1,029	500 (48.5)	63 (6.1)

Author	Year (data collection)	Country	Age (years)	Study Population (n)	Urban n (% of total)	Dementia n (% of total)
Hendrie 2018 ^{*18}	1992-2009	USA	≥70	3,259	2326 (71.4)	285 (8.7)
Huang 2014 ⁴⁵	1997-2007	Taiwan	Any (mean 58.7 +/- 14)	142,744	41,458 (29.0)	612 (0.43)
Jia 2014 ^{38*}	2008-2009	China	≥65	10,276	6,096 (59.3)	528 (5.1)
Khedr 2015 ^{6*}	2011-2013	Egypt	≥60	691		35 (5.1)
Kim 2011 ^{*39}	2008	South Korea	≥65	1,673	N/A	351 (21.0)
Koller 2010 ^{15*}	2004-2006	Germany	≥65	9,216	6,578 (71.4)	1846 (20.0)
Nadel 2014 ⁴⁰	2011	Costa Rica	>60 (mean 76.8+8.1)	90	N/A	2 (2.3)
Nunes 2010 ^{*34}	2003	Portugal	55-79	1,146	713 (62.2)	31 (2.7)
		Central African republic	≥65	973	452 (46.4)	72 (7.5)
Pilleron 2015 ¹³ (a)*	2011-2012					

Author	Year (data collection)	Country	Age (years)	Study Population (n)	Urban n (% of total)	Dementia n (% of total)
		Republic of				
Pilleron 2015 ¹³ (b)*	2011-2012	Congo	≥65	1,029	479 (46.6)	63(6.1)
Poddar 2011 ^{12*}	Unclear	India	≥50	2890	793 (27.4)	147 (5.1)
Prince 2012 ^{22*} (a)	2003-2010	Peru	>=65 (mean age follow-up 75.7- 78.6)	1933	1381	10/66: 69 (3.6) DSM-IV: 53 (2.7)
Prince 2012 ^{22*} (b)	2003-2010	Mexico	>=65 (mean age follow-up 75.7- 78.6)	2003	1003	10/66: 121 (6.0) DSM-IV: 67 (3.3)

Author	Year (data collection)	Country	Age (years)	Study Population (n)	Urban n (% of total)	Dementia n (% of total)
			>=65 (mean age follow-up 75.7-			10/66: 161 (7.4)
Prince 2012 ^{22*} (c)	2003-2010	China	78.6)	2162	1160	DSM-IV: 106 (4.9)
						2
Raina 2014 ^{7*}	unclear	India	>60	1000	500 (50)	3 (2.3)
Rodríguez-Agudelo						
2011 ^{23*}	2011	Mexico	>65	2003	1003 (50.1)	171 (8.5)
						After 2 years follow
		The Republic				up, dementia =23
Samba 2015 ^{9*}	2012-2014	of Congo	≥65	966	467 (48.3)	(2.38)

Author	Year (data collection)	Country	Age (years)	Study Population (n)	Urban n (% of total)	Dementia n (% of total)
			>=60 (mean age 69.2 years, 68.7-69.7)	1257	681 (54.2)	99 (7.9)
Sharifi 2016 ^{35*}	2012	Iran				
Taylor 2017 ⁴⁶	1999-2014	USA	Any age	Unclear	Unclear	1999= 44,536 2014 = 93,541
			>=60 (average age 70.5, SD = 7.22)	2980	1635 (54.9)	418 (14.0)
Tengku Aizan 2010 ^{10*}	2004-2005	Malaysia				
			>=65 (mean age 76.5 +/- 7.8)	2170	1459 (67.2)	184 (8.5)
Tola-Arribas 2013 ^{11*}	2009-2010	Spain				
Tornau 2015 ²⁰	Unknown (published 2015 as abstract)	Germany	>59	Unclear	Not specified	Not specified

Author	Year (data collection)	Country	Age (years)	Study Population (n)	Urban n (% of total)	Dementia n (% of total)
Tripathi 2012 ^{36*}	2003-2005	India	44-85	300	190 (63.3)	150 (50)
Weden 2018 ³⁷ (a)*	2000	USA	>=55	16,386	12,096 (73.8)	958 (5.8)
Weden 2018 ³⁷ (b)*	2010	USA	>=55	16,311	12,410 (76.1)	727 (4.5)
Wu 2017 ^{16*}	2008-2011	England	>=65	7505	5951 (79.3)	328 (4.4)
Yin 2016 ⁴⁷	2006-2012	China	>65	73 million approx.	Unclear	Unclear
Yuan 2016 ^{8*}	1997-2002	China	>=55	12,881	5687 (44.2)	311 (2.4)

Author	Year (data collection)	Country	Age (years)	Study Population (n)	Urban n (% of total)	Dementia n (% of total)
Zhao 2010 ⁴⁸	1997-1998	China	>=55	1235 completed phase 3	Unclear	475 (38.5)

AD = Alzheimer's Disease; ADRD = Alzheimer's Related Dementia; N/A = not available (unable to extract). Grey = not in meta-analysis

Additional information: Studies separating geographical areas e.g., “100% urban” and “mixed urban”, multiple different metropolitan sizes or conurbation and urban, were combined together as “urban”. One study³⁵ used rural areas as the referent, and we reversed the results using natural logarithms. One study¹⁸ had two cohorts but only raw data from authors available for one cohort to calculate Odds ratios.

Quantitative synthesis

Overall, rural living was associated with small increased odds of dementia, compared to urban living (OR, 1.20, 95% CI 1.03-1.40, P value = 0.0182; $N=25,274,741$, $N_{\text{Dementia}}=3,462,039$; 33 studies), **Figure 2**. However, stratifying by HICs and L-MICs status revealed a difference, with HICs studies showing no effect of rural vs urban (1.01, 0.88 to 1.16, P value = 0.8842; $N=25,221,600$, $N_{\text{Dementia}}=3,458,633$; 13 studies) and L-MICs suggesting increased odds of dementia in rural areas (1.33, 1.06 -1.68, P value = 0.0141; $N=53,141$, $N_{\text{Dementia}}=3406$ 20 studies). There was high heterogeneity: I^2 value = 99.78% (for all studies), 99.72% (for only HICs studies) and 87.64% (for only L-MICs studies).

Sensitivity analysis to explore the impact of potential overlap in L-MICs study participants^{9,13,22,23,26} did not significantly affect the results: rural living was still linked with a slight increase in dementia risk (1.35, 1.02-1.79, P value = 0.0354, I^2 = 90.69%). Including studies from the previous systematic review also showed a slight increased risk of dementia in rural areas, (1.18, 1.01-1.37) (**Appendix 6**).

Further stratifying HICs and L-MICs by study size confirmed in larger L-MICs studies ($N>10,000$) higher odds of dementia in rural areas (or lower in urban areas; 1.56, 1.20-2.02; $N=23,157$, $N_{\text{Dementia}}=839$; 2 studies), unchanged when moderate-sized studies ($N>1000$) were included (1.50, 1.017-1.93; $N=48,439$, $N_{\text{Dementia}}=2856$; 14 studies). However, larger HICs studies ($N>1,000,000$) showed *lower* odds of dementia in rural areas (0.88, 0.84-0.92; $N=24,900,843$, $N_{\text{Dementia}}=3,450,748$; 3 studies). This effect was reduced when moderate-sized studies ($N>10,000$) were included (0.93, 0.76-1.13; $N=25,195,209$, $N_{\text{Dementia}}=3,455,479$; six studies). Studies in L-MICs were generally smaller than those in HICs.

Study	Country	Total N	Dementia	OR (95% CI)
Goodman 2017	USA	21,624,228	3,110,654	0.84 (0.84, 0.85)
Chen 2017	Canada	2,066,639	257,816	0.91 (0.90, 0.92)
Abner 2016	USA	1,209,976	82,278	0.89 (0.88, 0.91)
High Income Countries: N>1,000,000				0.88 (0.84, 0.92)
Astell-Burt 2018	Australia	261,669	3,046	0.63 (0.56, 0.73)
Weden 2018a	USA	16,386	958	1.34 (1.10, 1.63)
Weden 2018b	USA	16,311	727	1.17 (0.92, 1.48)
High Income Countries: N>10,000				0.93 (0.76, 1.13)
Koller 2010	Germany	9,216	1,846	0.96 (0.86, 1.08)
Wu 2017	England	7,505	328	0.92 (0.64, 1.33)
Hendrie 2018	USA	3,276	285	1.56 (1.21, 2.01)
Tola-Arribas 2013	Spain	2,170	184	0.99 (0.72, 1.37)
Kim 2011	South Korea	1,673	351	1.17 (1.00, 1.38)
Goncalves-Pereira 2017	Portugal	1,405	129	0.99 (0.69, 1.42)
Nunes 2010	Portugal	1,146	31	2.12 (0.91, 4.96)
All High Income Countries				1.01 (0.88, 1.16)
Yuan 2016	China	12,881	311	1.32 (1.05, 1.66)
Jia 2014	China	10,276	528	1.45 (1.22, 1.73)
Low-to-Middle Income Countries: N>10,000				1.40 (1.22, 1.61)
Chen 2012b	China	3,327	341	2.26 (1.79, 2.86)
Tengku Aizan 2010	Malaysia	2,980	418	1.87 (1.52, 2.31)
Chen 2012a	China	2,917	210	4.27 (3.13, 5.84)
Poddar 2011	India	2,890	146	1.49 (0.99, 2.24)
Prince 2012c	China	2,162	161	0.93 (0.67, 1.29)
Prince 2012b	Mexico	2,003	121	2.24 (1.51, 3.31)
Rodríguez-Agudelo 2011	Mexico	2,003	171	0.99 (0.72, 1.36)
Prince 2012a	Peru	1,933	69	1.36 (0.82, 2.23)
Deng 2018	China	1,781	195	1.56 (0.99, 2.45)
Sharifi 2016	Iran	1,257	99	1.88 (1.23, 2.86)
Guerchet 2013b	ROC	1,029	63	0.85 (0.51, 1.42)
Raina 2014	India	1,000	23	0.43 (0.18, 1.05)
Low-to-Middle Income Countries: N>1000				1.50 (1.17, 1.93)
Guerchet 2013a	CAR	973	72	1.35 (0.83, 2.19)
Samba 2015	ROC	966	23	1.78 (0.75, 4.23)
Pilleron 2015b	ROC	912	135	1.01 (0.60, 1.68)
Pilleron 2015a	CAR	860	135	1.43 (0.88, 2.32)
Khedr 2015	Egypt	691	35	0.44 (0.22, 0.90)
Tripathi 2012	India	300	150	0.56 (0.35, 0.90)
All Low-to-Middle Income Countries				1.33 (1.06, 1.68)
All Studies				1.20 (1.03, 1.40)

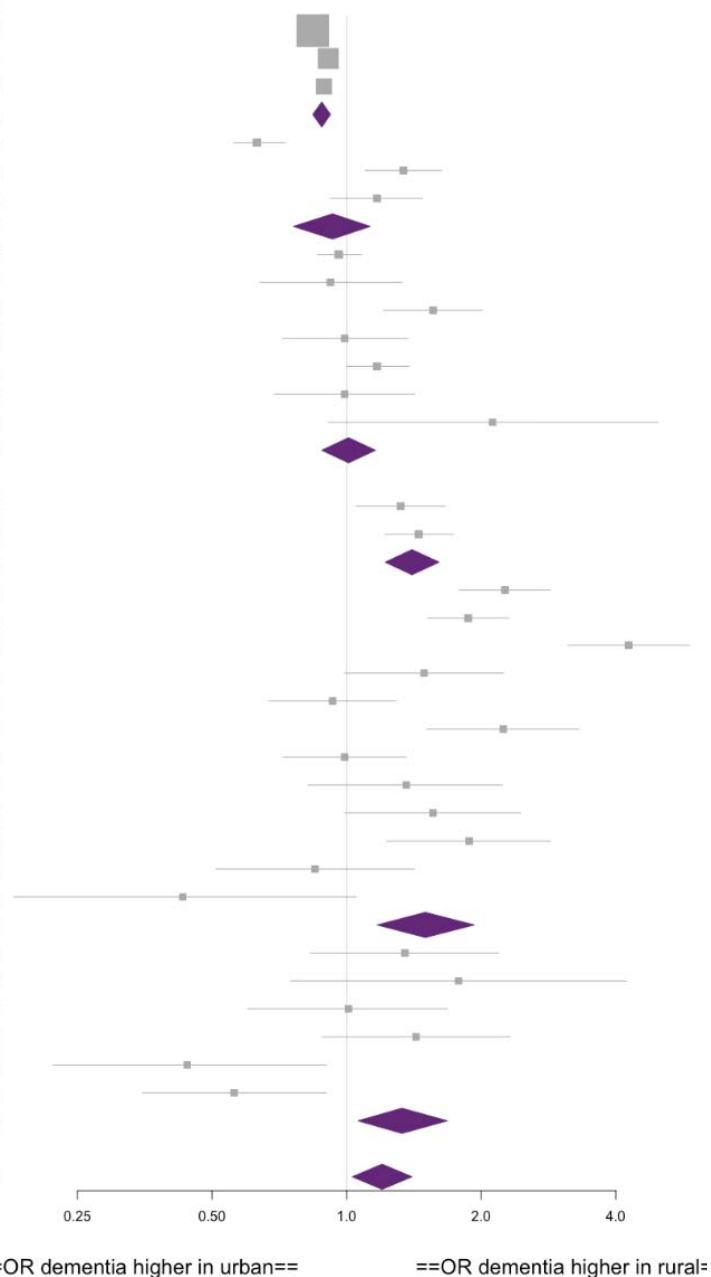


Figure 2: Forest plot showing meta-analysed rural/urban odds ratios of dementia stratified by HICs/L-MICs status and size of study: rural/urban differences in dementia ¹.

¹.Guerchet 2013, Pilleron 2015, Prince 2012 (Mexico only) Rodríguez-Agudelo 2011 and Samba 2015: unclear if there is overlap in these L-MICs papers (as more than one similar study in the same country). Sensitivity analysis removing the smaller cohorts (Pilleron 2015, Rodríguez-Agudelo 2011 and Samba 2015) shows no significant change in result

Narrative synthesis of non-meta-analysis studies

The 11 heterogeneous papers not included in the meta-analysis, were undertaken in Europe (Switzerland, Romania, Germany and Spain),^{20,41-43} South-America (Brazil),⁴⁴ Central America (Costa Rica),⁴⁰ North America (USA)^{17,46} and Asia (Taiwan and China).^{45,47,48} between 1993 and 2012. One study investigated dementia in diabetic patients,⁴⁵ the other studies investigated dementia in general. Five studies had no age restriction,^{17,42,43,45,46} the others included people >55/59 years or >60/65 years.^{20,40,41,44,47,48}, with sample size from 90 to approximately 73 million (median, 17,186).

The dementia ascertainment methodologies varied greatly: data linkage with ICD codes;^{17,20,42,43,45-47} cross sectional incidence surveys; questionnaires; structured interviews and/or screening by clinical assessment.^{40,41,44,48}

Just over half of these studies reflected the findings of the meta-analysis, with an increase of dementia in rural areas.^{41,43,44,46-48} The results were unclear in several of these studies,^{20,40} and showed no difference in one study investigating the variation of AD drug dispensing¹⁷. Of the two studies suggesting an increase in urban areas, one was specifically investigating dementia (AD) deaths,⁴² and the other newly diagnosed diabetic patients.⁴⁵

Most studies only assessed the relationship between residence at one time point and subsequent dementia, but one study⁴¹ investigated childhood and adulthood residence, finding childhood rural residence was linked with a non-significant trend for risk of dementia (HR = 1.52, 95% CI = 0.93-2.49, p=0.08), which became significant in adulthood (HR 1.61, 95% CI 1.06-2.46, p<0.05).⁴¹, but there was no association in late-life (HR 0.78, 95% CI 0.49-1.23, p=0.29).⁴¹

Risk of Bias

Figures 3 and **Appendix 7** show the risk of bias assessments for included studies. There was great variation in the risk of bias and, in particular, high risk in measurement of exposure (geographical location) in approximately half of the studies.

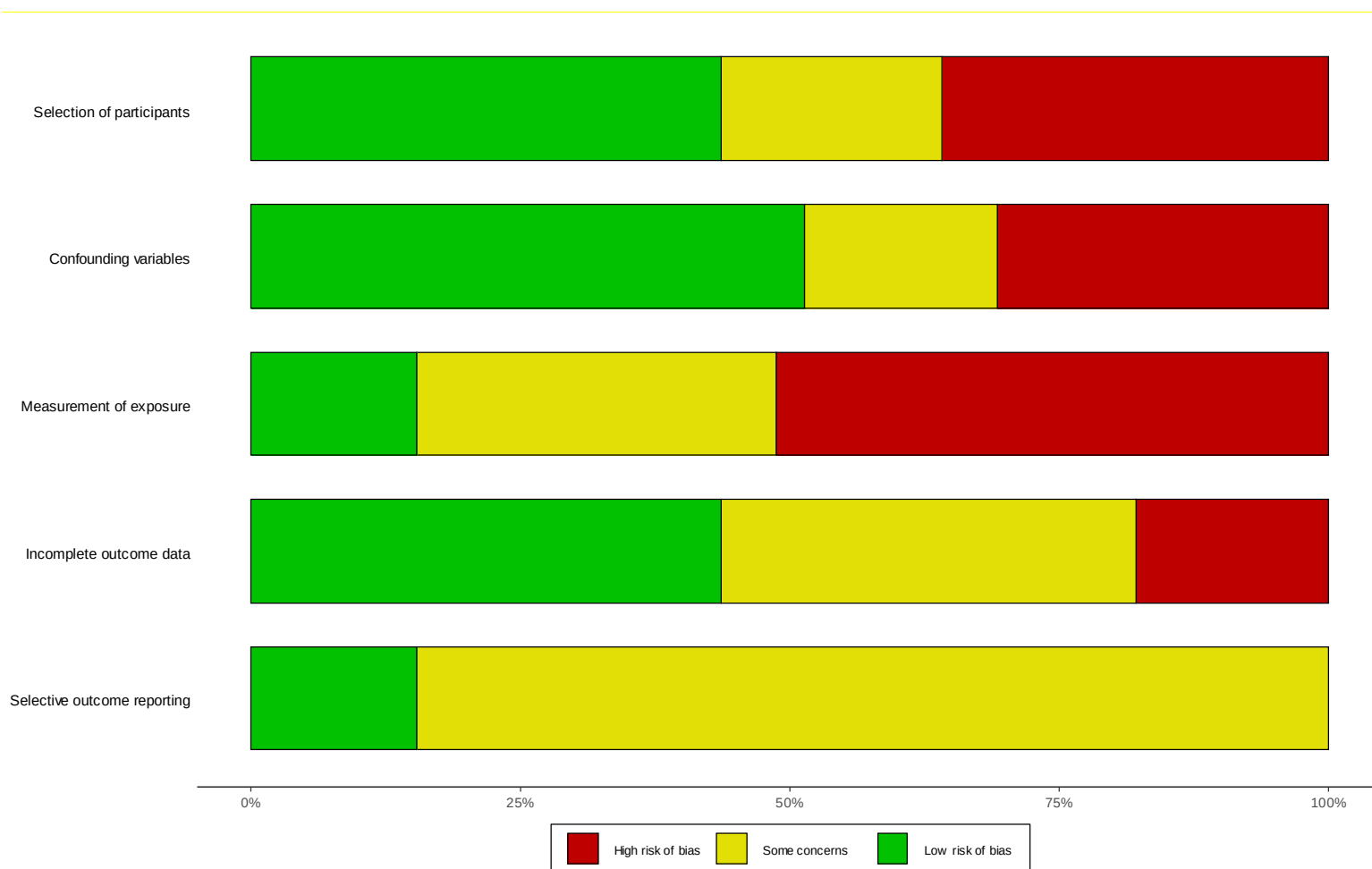


Figure 3. Summary of Risk of Bias assessment: rural/urban differences in dementia

DISCUSSION

Overall, rural living was associated with small increased odds of dementia compared to urban living (OR, 1.20, 95% CI 1.03-1.40, P value = 0.0182; $N=25,274,741$, $N_{\text{Dementia}}=3,462,039$; 33 studies). Further stratification by HICs and L-MICs showed this difference was mostly driven by L-MICs (1.33, 1.06 -1.68, P value = 0.0141; $N=53,141$, $N_{\text{Dementia}}=3406$ 20 studies). In large HICs, odds of dementia were higher in urban areas (0.88, 0.84-0.92). To our knowledge, this difference has not been reported before, and has important implications for research and policy in this area. Studies were heterogeneous in the definition of ‘rural’ or ‘urban’ and identification of dementia, and there was high risk of bias in the exposure (geographical variation). Studies often did not report when in the life-course the geographical exposure was measured, and whether this changed across their lifetime, and the association between rurality and dementia prevalence and incidence may change with at different ages.

Increased risk of dementia in those living in rural areas may be due to specific aspects of living in these geographical locations, such as access to healthcare services/dementia diagnosis, air pollution or social connections, or may be confounded or mediated by education or socio-economic status.¹ There is a complex interplay between the environment and the individual living in it. For dementia risk, this includes direct and indirect associations between the environment, and cognition, and noting that both of these can change across the life course.⁴⁹

Dementia prevalence can vary country to country, influenced by the drive to seek a dementia diagnosis as well as the different the methodologies used for dementia ascertainment. A recent systematic review recommends using multiple sources, including access to medical records, to ascertain dementia to improve consistency in methodologies between studies.^{50,51}

For example, in another systematic review⁵² which found a mixed picture between green space exposure and dementia risk, green space exposure was associated with higher incidence of dementia measured by anti-dementia medication and reversed if the other methods for detecting dementia were used.⁵³ It is therefore essential that the method(s) of dementia ascertainment is clearly reported and the potential limitations and biases of this acknowledged.

When comparing multiple studies it is important that the dementia ascertainment rate for the study population is calculated and compared with an expected rate.⁵¹ Dementia diagnosis, particularly in L-MICs, can still be challenging, and there is ongoing debate about the best criteria for dementia diagnosis:⁵⁴ criteria (such as the 10/66 criteria) may have greater accuracy than DSM-IV in low levels of literacy.⁵⁴ Furthermore, in rural areas in general, or L-MICs, there may be limited access to diagnostic assessment,⁵⁵ especially if services are centralised.⁵⁶ Our systematic review did not show an increased risk of dementia in urban settings to suggest this bias, but this needs considered when data are included from more countries.

There may be increased cognitive reserve in people living in urban areas^{49,57,58}, particularly during childhood, which influences dementia variation between rural and urban living. The impact from social deprivation may be higher than that from rural or urban location (e.g. the odds of a cognitive disorder were twice as high in those living in poor-quality rural environments, than those in rural areas).¹⁶

Education influences dementia risk,⁵⁹ but few studies investigated education in relation to the geographical area across time. Socio-economic status (SES) also influences dementia prevalence rates^{50,60}. Our systematic review was unable to assess for SES in influencing the difference in rural versus urban dementia risk. Whilst education, SES and other variables for

overall dementia risk were often reported, these variables were rarely considered in relation to rural or urban areas. Further work is required to understand the complex interplay between geographical location, education, socioeconomic circumstances, prior cognitive ability and dementia risk: whether there are critical time points in a person's life where individual or combined factors influence dementia risk. Future studies should include more detailed information on participants' education, SES measures in relation to geographical area, as well as the different definitions of geographical categorisation.

Comparison with other literature

This large and comprehensive systematic review builds on a previous systematic review in 2010 which showed, in 25 rural versus urban studies, that there was some evidence of greater all-cause dementia odds with rural living [prevalence OR = 1.11, 90% CI 0.79-1.57; incidence OR = 1.20, 90% CI 0.84-1.71], and that the association between rurality and dementia risk was stronger in studies only including people with Alzheimer Disease (OR=1.50, 90% CI 1.33-1.69), but weaker in studies only including people with vascular dementia (OR= 1.09, 90% CI 0.65-1.83). The review also suggested that exposure to rural living early in life further increases the risk of subsequent dementia.

In comparison, our subsequent review which included 38 new papers was unable to investigate the dementia sub-type risk, or the life course effect of geographical location, as too few studies which reported adequate data on these predictors to undertake meaningful meta-analysis. Most studies used dementia as a 'blanket' term encompassing different sub-types of dementia, and did not provide adequate data for the life course effect of geographical location on dementia risk. The previous review investigated dementia prevalence and incidence at a number of geographical scales (country-to-country comparisons, rural to urban comparisons,

and regional comparisons), but there has been a rapid increase studies published on this topic, therefore our current review focussed on rural to urban comparisons.

Limitations and strengths: of the systematic review

The literature search for this comprehensive systematic review was developed with an information specialist and used a validated comprehensive search strategy from our previous review, with a pre-registered protocol, and was reported according to PRISMA guidance. At least two reviewers independently reviewed titles and abstracts, and extracted and analysed the studies with high levels of agreement using proforma which was produced and trialled for data extraction, and risk of bias (adapting the ROBANS tool). Authors were contacted for additional information, to increase data available for meta-analysis. The language was not limited to English to minimise bias in selecting papers for review.

The limitations of this study include that despite a comprehensive search strategy including forward citation of included studies; it is possible that relevant papers may have been missed, as we did not search for grey literature. There is also significant variation within and between countries of the level of geography which is recorded.

The Risk of Bias assessment tool was based on ROBANS and whilst it did assess the quality of the method of dementia ascertainment, for the studies using big data and ICD-10 codes it was not possible to assess the quality of that dementia diagnosis.

Limitations and strengths: of the included studies

We identified 38 new relevant papers in the last 5 years, and were able to extract, or calculate, odds ratios for 27 papers (33 studies) for the meta-analysis, including studies from a range of

L-MICs and HICs. The sample sizes were overall large (range 90 to 73,000,000, median 2170 for all papers, and 300 to 21,624,228, median 2162, for those in the meta-analysis) and the method of defining dementia was reported in the studies.

The individual studies were heterogeneous in their methodologies. In particular, there was significant heterogeneity in of the definition of ‘rural’ and ‘urban’, and how this translates at an international level for comparison. Too few studies reported the impact of geographical location at different time points in a person’s life to include in meta-analysis. This means there could be confounding or bias influencing the risk of dementia.

There is a need for standard definitions of urbanisation and rurality, and a clearer definition of “place.”⁴⁹ Despite proposed quantifiable measures of urbanisation through population density and sprawl and further evaluation of physical characteristics of proximal environments,⁴⁹ none of our included papers considered these factors.

It is also important to acknowledge that diagnostic coverage, those diagnosed with dementia rather than the true dementia prevalence, can be low⁶¹ and therefore missed cases (particularly within studies using large big datasets of routinely collected data⁶²) may introduce bias, potentially under-reporting, in interpreting geographical variation of dementia versus traditional face-to-face epidemiological studies.

Implications

There has been a rapid increase in interest in the geographical predictors of dementia, therefore this study, limited only to urban versus rural geographical variation, confirmed previous findings of increased odds of subsequent dementia for people living in rural areas. The heterogeneity of the studies means understanding rural versus urban risk of dementia is complicated. To ensure a future robust study to investigate the geographical variation of rural

versus urban living and dementia risk (given our finding of OR of 1.16 to 1.26), the sample size would need to be approximately 1000-4 000 people.

Further consensus on the term ‘geographical location’ with clear definitions of ‘rural’ and ‘urban’ that is translatable across, and within, different countries is important to understand the complex interplay between the environment and the individual in relation to dementia risk. Future large-scale studies must also include detailed information about the area (including a measure of deprivation, greenspace, access to services etc.) and a person’s geographical location across time.

We know the incidence and prevalence of dementia is rising in L-MICs, but larger studies in this systematic review are in HICs, where they are more likely to have healthcare systems with routinely collected data allowing larger scale studies. Whereas, L-MICs are more likely to use traditional (smaller scale) face-to-face epidemiological studies. As these research methods are more costly than analysing nationally collected data, it results in fewer studies in L-MICs, with lower numbers of participants, but likely to provide more detailed analysis. Finally, the definition of LIC, HMIC and HIC has the potential to change with time, which is important to consider given the longitudinal aspect for dementia risk.

Conclusions

Overall, rural living was associated with small-increased odds of dementia, compared to urban living (OR, 1.20, 95% CI 1.03-1.40, P value = 0.0182). Stratifying by HICs and L-MICs suggested increased odds of dementia in L-MICs rural areas, (1.33, 1.06 -1.68, P value = 0.0141). However, methodological differences complicate combining the extant literature. Therefore, better-designed studies with careful consideration of mediating and confounding variables are required to answer this important question definitively.

Acknowledgements

KEW was supported by clinical research fellowship from Alzheimer Scotland and the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-council Lifelong Health and Wellbeing initiative (MR/L501530/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. KEW is also supported by Alzheimer Scotland Dementia Research Centre.

JKB is supported by an NHS Education for Scotland/Chief Scientists Office Postdoctoral Clinical Lectureship (PCL/21/01).

GMT is supported by the Osteopathic Heritage Foundation Ralph S. Licklider D.O.

Lastly, we would like to thank Professor John Starr (co-author) for his huge lifetime contribution to dementia research worldwide. We are deeply saddened by his sudden unexpected death on 9 December 2018 and miss our colleague greatly. His input as co-author was important to this research.

REFERENCES

1. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 2020;**396**(10248):413-446.
2. Russ TC, Batty GD, Hearnshaw GF, Fenton C, Starr JM. Geographical variation in dementia: systematic review with meta-analysis. *International Journal of Epidemiology* 2012;**41**(4):1012-1032.
3. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;**4**(1):1.
4. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj* 2021;**372**:n71.
5. Covidence: Systematic Review Management. <https://www.covidence.org> Accessed 14th September 2023.
6. Khedr E, Fawi G, Abbas MA, et al. Prevalence of mild cognitive impairment and dementia among the elderly population of Qena Governorate, Upper Egypt: a community-based study. *Journal of Alzheimer's Disease* 2015;**45**(1):117-26.
7. Raina SK, Raina S, Chander V, et al. Is dementia differentially distributed? A study on the prevalence of dementia in migrant, urban, rural, and tribal elderly population of Himalayan region in Northern India. *North American Journal of Medical Sciences* 2014;**6**(4):172-177.
8. Yuan J, Zhang Z, Wen H, et al. Incidence of dementia and subtypes: A cohort study in four regions in China. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2016;**12**(3):262-271.
9. Samba H, Guerchet M, Bandzouzi BN, et al. Incidence of dementia among older adults in central africa: First results from the republic of congo in the EPIDEMCA-FU study. *Alzheimer's and Dementia* 2015;**1**:P221-P222.
10. Tengku Aizan H, Krishnaswamy S, Siti Suhailah A, Yadollah Abolfathi M. Sociodemographic risk factors and correlates of dementia in older Malaysians. *Dementia and Geriatric Cognitive Disorders* 2010;**30**(6):533-539.

11. Tola-Arribas MA, Yugueros MI, Gareia MJ, et al. Prevalence of dementia and subtypes in Valladolid, northwestern Spain: The DEMIN Vall Study. *PLoS ONE* 2013;**8**(10).
12. Poddar K, Kant S, Singh A, Singh TB. An epidemiological study of dementia among the habitants of Eastern Uttar Pradesh, India. *Annals of the Indian Academy of Neurology* 2011;**14**(3):164-168.
13. Pilleron S, Clement JP, Ndamba-Bandzouzi B, et al. Is dependent personality disorder associated with mild cognitive impairment and dementia in Central Africa? A result from the EPIDEMCA programme. *International Psychogeriatrics* 2015;**27**(2):279-288.
14. Kim KW, Park JH, Kim MH, et al. A nationwide survey on the prevalence of dementia and mild cognitive impairment in South Korea. *Journal of Alzheimer's Disease* 2011;**23**(2):281-291.
15. Koller D, Eisele M, Kaduszkiewicz H, et al. Ambulatory health services utilization in patients with dementia - is there an urban-rural difference? *International Journal of Health Geographics* 2010;**9**(59).
16. Wu YT, Prina AM, Jones A, et al. The Built Environment and Cognitive Disorders: Results From the Cognitive Function and Ageing Study II. *American Journal of Preventive Medicine* 2017;**53**(1):25-32.
17. Ganesh C. The diffusion of prescription drugs for Alzheimer's disease among medicare beneficiaries [Ph.D.]. The Pennsylvania State University, 2010.
18. Hendrie HC, Smith-Gamble V, Lane KA, et al. The Association of Early Life Factors and Declining Incidence Rates of Dementia in an Elderly Population of African Americans. *Journals of Gerontology Series B: Psychological Sciences & Social Sciences* 2018;**73**:S82-S89.
19. Guerchet M, Bandzouzi-Ndamba B, Mbelesso P, et al. Prevalence of dementia in two countries of central africa: Comparison of rural and urban areas in the epidemca study. *Neuroepidemiology* 2013;**41** (3-4):253-254.
20. Tornau F, Bruchmann G, Juckel G. Prevalence of psychiatric disorders in the elderly in rural and urban areas in Germany. *International Psychogeriatrics* 2015;**27**:S167.

21. Goncalves-Pereira M, Cardoso A, Verdelho A, et al. The prevalence of dementia in a Portuguese community sample: a 10/66 Dementia Research Group study. *BMC Geriatrics* 2017;**17**(1):261.
22. Prince M, Acosta D, Ferri CP, et al. Dementia incidence and mortality in middle-income countries, and associations with indicators of cognitive reserve: A 10/66 Dementia Research Group population-based cohort study. *The Lancet* 2012;**380**(9836):50-58.
23. Rodriguez-Agudelo Y, Solis-Vivanco R, Acosta-Castillo I, et al. Neuropsychiatric symptoms in older adults with and without dementia in urban and rural regions. Results of the 10/66 Dementia Research Group in Mexico. [Spanish]

Sintomas neuropsiquiatricos en adultos mayores con y sin demencia de regiones urbana y rural. Resultados del Grupo de Investigacion en Demencia 10/66 en Mexico. *Revista de Investigacion Clinica* 2011;**63**(4):382-390.
24. Prince M. Dementia in developing countries. A consensus statement from the 10/66 Dementia Research Group. *Int J Geriatr Psychiatry* 2000;**15**(1):14-20.
25. Prina AM, Mayston R, Wu YT, Prince M. A review of the 10/66 dementia research group. *Soc Psychiatry Psychiatr Epidemiol* 2019;**54**(1):1-10.
26. Guerchet M, Ndamba-Bandzouzi B, Mbelesso P, et al. Comparison of rural and urban dementia prevalences in two countries of central africa: The epidemca study. *Alzheimer's and Dementia* 2013;**1**:P688.
27. Wellcome. Low-and middle-income countries. <https://wellcome.org/grant-funding/guidance/low-and-middle-income-countries> Accessed 8th September, 2023.
28. Abner EL, Jicha GA, Christian WJ, Schreurs BG. Rural-urban differences in Alzheimer's disease and related disorders diagnostic prevalence in Kentucky and West Virginia. *Journal of Rural Health* 2016;**32**(3):314-320.
29. Astell-Burt T, Feng X. Is the risk of developing Alzheimer's disease really higher in rural areas? A multilevel longitudinal study of 261,669 Australians aged 45 years and older tracked over 11 years. *Health and Place* 2018;**54**:132-137.

30. Chen H, Kwong JC, Copes R, et al. Exposure to ambient air pollution and the incidence of dementia: A population-based cohort study. *Environment International* 2017;**108**:271-277.
31. Chen R, Ma Y, Wilson K, et al. A multicentre community-based study of dementia cases and subcases in older people in China--the GMS-AGECAT prevalence and socio-economic correlates. *International Journal of Geriatric Psychiatry* 2012;**27**(7):692-702.
32. Deng J, Cao C, Jiang Y, et al. Prevalence and effect factors of dementia among the community elderly in Chongqing, China. *Psychogeriatrics* 2018;**18**(5):412-420.
33. Goodman RA, Lochner KA, Thambisetty M, et al. Prevalence of dementia subtypes in United States Medicare fee-for-service beneficiaries, 2011-2013. *Alzheimer's & Dementia* 2017;**13**(1):28-37.
34. Nunes B, Silva RD, Cruz VT, et al. Prevalence and pattern of cognitive impairment in rural and urban populations from Northern Portugal. *BMC Neurology Vol 10 2010, ArtID 42* 2010;**10**.
35. Sharifi F, Fakhrzadeh H, Varmaghani M, et al. Prevalence of Dementia and Associated Factors among Older Adults in Iran: National Elderly Health Survey (NEHS). *Archives of Iranian Medicine* 2016;**19**(12):838-844.
36. Tripathi M, Vibha D, Gupta P, et al. Risk factors of dementia in North India: A case-control study. *Aging & Mental Health* 2012;**16**(2):228-235.
37. Weden MM, Shih RA, Kabeto MU, Langa KM. Secular trends in dementia and cognitive impairment of U.S. rural and urban older adults. *American Journal of Preventive Medicine* 2018;**54**(2):164-172.
38. Jia J, Wang F, Wei C, et al. The prevalence of dementia in urban and rural areas of China. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2014;**10**(1):1-9.
39. Kim K, Park J, Kim M, et al. A nationwide survey on the prevalence of dementia and mild cognitive impairment in South Korea. *JAD, Journal of Alzheimer's Disease* 2011;**23**(2):281-291.
40. Nadel JL, Ulate D. Incidence and risk factors for cognitive impairment in rural elderly populations in Costa Rica^ien

Incidencia y factores de riesgo para la discapacidad cognitiva en poblaciones rurales de tercera edad en Costa Rica^{ies}. *Rev. biol. trop* 2014;**62**(3):869-876.

41. Contador I, Bermejo-Pareja F, Puertas-Martin V, Benito-Leon J. Childhood and Adulthood Rural Residence Increases the Risk of Dementia: NEDICES Study. *Current Alzheimer Research* 2015;**12**(4):350-7.
42. Chammartin F, Probst-Hensch N, Utzinger J, Vounatsou P. Mortality atlas of the main causes of death in Switzerland, 2008-2012. *Swiss Medical Weekly* 2016;**146**:w14280.
43. Cornutiu G. The incidence and prevalence of Alzheimer's disease. *Neurodegenerative Diseases* 2010;**8**(1-2):9-14.
44. Silva EFd, Paniz VMV, Laste G, Torres ILdS. Prevalência de morbidades e sintomas em idosos: um estudo comparativo entre zonas rural e urbana. *Ciência & Saúde Coletiva* 2013;**18**(4):1029-1040.
45. Huang C, Chung C, Leu H, et al. Diabetes mellitus and the risk of Alzheimer's disease: a nationwide population-based study. *PLoS ONE* 2014;**9**(1).
46. Taylor CA, Greenlund SF, McGuire LC, Lu H, Croft JB. Deaths from Alzheimer's Disease - United States, 1999-2014. *Mmwr* 2017;**Morbidity and mortality weekly report**. **66**(20):521-526.
47. Yin P, Feng X, Astell-Burt T, et al. Temporal trends and geographic variations in dementia mortality in China between 2006 and 2012: Multilevel evidence from a nationally representative sample. *Alzheimer Disease and Associated Disorders* 2016;**30**(4):348-353.
48. Zhao Q, Zhou B, Ding D, Guo Q, Hong Z. Prevalence, mortality, and predictive factors on survival of dementia in Shanghai, China. *Alzheimer Disease and Associated Disorders* 2010;**24**(2):151-158.
49. Cassarino M, Setti A. Environment as 'Brain Training': A review of geographical and physical environmental influences on cognitive ageing. *Ageing Research Reviews* 2015;**Part B**. **23**:167-182.
50. Mullin DS, Stirland LE, Buchanan E, et al. Identifying dementia using medical data linkage in a longitudinal cohort study: Lothian Birth Cohort 1936. *BMC Psychiatry* 2023;**23**(1):303.

51. Sibbett RA, Russ TC, Deary IJ, Starr JM. Dementia ascertainment using existing data in UK longitudinal and cohort studies: a systematic review of methodology. *BMC Psychiatry* 2017;**17**(1):239.
52. Chen X, Lee C, Huang H. Neighborhood built environment associated with cognition and dementia risk among older adults: A systematic literature review. *Soc Sci Med* 2022;**292**:114560.
53. Astell-Burt T, Navakatikyan MA, Feng X. Urban green space, tree canopy and 11-year risk of dementia in a cohort of 109,688 Australians. *Environ Int* 2020;**145**:106102.
54. Ferri CP, Oliveira D. Harmonization of epidemiological studies on dementia in Latin America Why does it matter? *Dement Neuropsychol* 2019;**13**(4):363-366.
55. Zakarias JK, Jensen-Dahm C, Nørgaard A, et al. Geographical Variation in the Diagnostic Rate and Quality of Dementia Diagnoses. *J Alzheimers Dis* 2019;**69**(2):513-520.
56. Jorgensen TS, Torp-Pedersen C, Gislason GH, Andersson C, Holm E. Time trend in Alzheimer diagnoses and the association between distance to an Alzheimer clinic and Alzheimer diagnosis. *European journal of public health* 2015;**25**(3):522-527.
57. Brayne C, Wu YT. Population-Based Studies in Dementia and Ageing Research: A Local and National Experience in Cambridgeshire and the UK. *Am J Alzheimers Dis Other Demen* 2022;**37**:15333175221104347.
58. Cassarino M, O'Sullivan V, Kenny RA, Setti A. Environment and cognitive aging: A cross-sectional study of place of residence and cognitive performance in the Irish longitudinal study on aging. *Neuropsychology* 2016;**30**(5):543-557.
59. Cadar D, Lassale C, Davies H, et al. Individual and area-based socioeconomic factors associated with dementia incidence in England: Evidence from a 12-year follow-up in the English longitudinal study of ageing. *JAMA Psychiatry* 2018;**75**(7):723-732.
60. Matthews FE, Arthur A, Barnes LE, et al. A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: results of the Cognitive Function and Ageing Study I and II. *Lancet* 2013;**382**(9902):1405-12.

61. Gauthier S, Rosa-Neto P, Morais JA, C W. World Alzheimer Report 2021: Journey through the diagnosis of dementia

London, England: Alzheimer's Disease International, 2021.

62. Wilkinson T, Schnier C, Bush K, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol* 2019;**34**(6):557-565.

2.3.1 Systematic review appendices

The following Appendices 1-7 are from the manuscript uploaded to medRxiv and included here in the body of the Chapter for continuity. Appendix 1 includes the databases and search strategies used for the systematic review. Appendix 2 highlights the eligibility criteria. Appendix 3 outlines the country income definitions according to the World Bank in the year and how countries were assigned to High Income or Low-Middle Income Countries for the analysis. Appendix 4 is the adapted ROBINS quality assessment form used for the review. Appendix 5 is a supplementary table with additional information extracted from the studies in the systematic review (with those studies not included in the meta-analysis highlighted in grey). Appendix 6 is a supplementary forest plot figure including the Odds Ratios from the previous systematic review in 2012 for rural/urban variation and my systematic review Odds Ratios. Appendix 7 is a supplementary figure of all the included studies and risk of bias.

Appendix 1: Databases and search strategies

Databases searched: rural/urban differences in dementia

Database	Database start	Search date
ASSIA	1987	18.10.18
MEDLINE	1950	18.10.18
EMBASE	1974	18.10.18
CINAHL	1981	18.10.18
GeoRef	1669	18.10.18
Global Health	1973	18.10.18
LILACS	1982	23.10.18
PsycINFO	1806	18.10.18
COPAC	1996	23.10.18
SciELO	1997	23.10.18
Dissertations and Theses	1861	31.10.18
PapersFirst	1993	23.10.18
ProceedingsFirst	1993	23.10.18

The following databases used in our previous systematic review on this topic are no longer active and therefore were not included: FRANCIS, GEOBASE, Australian Digital Theses Program, ADT, Theses Canada Portal, Conference Papers Index

Indicative search strategies for Ovid Medline, EMBASE, and ASSIA: rural/urban differences in dementia

Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

1	Dementia/ep
2	Alzheimer Disease/ep
3	Delirium/ep
4	Wernicke Encephalopathy/ep
5	Delirium, Dementia, Amnestic, Cognitive Disorders/ep
6	dement*.ti,ab.
7	alzheimer*.ti,ab.
8	(lewy* adj2 bod*).ti,ab.
9	deliri*.ti,ab.
10	(chronic adj2 cerebrovascular).ti,ab.
11	"organic brain disease".ti,ab.
12	"organic brain syndrome".ti,ab.

13	("normal pressure hydrocephalus" and "shunt*").ti,ab.
14	"benign senescent forgetfulness".ti,ab.
15	(cerebr* adj2 deteriorat*).ti,ab.
16	(cerebr* adj2 deteriorat*).ti,ab.
17	(pick* adj2 disease).ti,ab.
18	huntington*.ti,ab.
19	binswanger*.ti,ab.
20	korsako*.ti,ab.
21	or/1-20
22	Incidence/
23	Prevalence/
24	Epidemiologic Studies/
25	Incidence.ti,ab.
26	Prevalence.ti,ab.
27	distribut*.ti,ab.
28	epidem*.ti,ab.
29	"geographical variation*".ti,ab.
30	or/22-29
31	Geography/
32	North America/ep
33	Pacific States/ep
34	Asia/ep
35	Africa/ep
36	Canada/ep
37	Europe/ep
38	Far East/ep
39	Scotland/ep
40	United States/ep
41	Sweden/ep
42	Spain/ep
43	(Geograph* or cluster* or AFGHANISTAN or "ÅLAND ISLANDS" or ALBANIA or SAMOA or ANDORRA or ANGOLA or ANGUILLA or ANTARCTICA or ANTIGUA or BARBUDA or ARGENTINA or ARMENIA or ARUBA or AUSTRALIA or AZERBAIJAN or BAHAMAS or BANGLADESH or BARBADOS or BELGIUM or BELIZE or BERMUDA or BOLIVIA or BONAIRE or BOSNIA).tw.
44	(HERZEGOVINA or BOTSWANA or "BOUVET ISLAND" or BRAZIL or "BRITISH INDIAN OCEAN TERRITORY" or BRUNEI or DARUSSALAM or BULGARIA or "BURKINA FASO" or BURUNDI or CAMBODIA or CAMEROON or CANADA or CAPE VERDE or "CAYMAN ISLANDS" or "CENTRAL AFRICAN REPUBLIC" or CHAD or CHILE or CHINA or "CHRISTMAS ISLAND").tw.
45	("COCOS ISLANDS" or "KEELING islands" or COLOMBIA or COMOROS or CONGO or "COOK ISLANDS" or "COSTA CÔTE" or CROATIA or CUBA or CURACAO or CYPRUS or CZECH* or DENMARK or DJIBOUTI or DOMINICA or "DOMINICAN REPUBLIC" or ECUADOR or EGYPT or "EL SALVADOR").tw.
46	("EQUATORIAL GUINEA" or ERITREA or ESTONIA or ETHIOPIA or "FALKLAND ISLANDS" or "FAROE ISLANDS" or FIJI or FINLAND or FRANCE or GUIANA or

	POLYNESIA or "FRENCH SOUTHERN TERRITORIES" or GABON or GAMBIA or GEORGIA or GERMANY or GHANA or GIBRALTAR or GREECE).tw.
47	(GREENLAND or GRENADA or GUADELOUPE or GUAM or GUATEMALA or GUERNSEY or GUINEA or "GUINEA-BISSAU" or HAITI or "HOLY SEE" or HONDURAS or "HONG KONG" or HUNGARY or ICELAND or INDIA or INDONESIA or IRAN or IRAQ or IRELAND or "ISLE OF MAN" or ISRAEL or ITALY or JAMAICA or JAPAN or JERSEY or JORDAN or KAZAKHSTAN or KENYA).tw.
48	(KIRIBATI or KOREA or KUWAIT or KYRGYZSTAN or LAO or LATVIA or LEBANON or LESOTHO or LIBERIA or LIBYA or LIECHTENSTEIN or LITHUANIA or LUXEMBOURG or MACAO or "MO MAC" or MACEDONIA or MADAGASCAR or MALAWI or MALAYSIA or MALDIVES or MALI or MALTA).tw.
49	("MARSHALL ISLANDS" or MARTINIQUE or MAURITANIA or MAYOTTE or MEXICO or MICRONESIA or MOLDOVA or MONACO or MONGOLIA or MONTserrat or MOROCCO or MOZAMBIQUE or MYANMAR or NAMIBIA or NAURU or NEPAL or NETHERLANDS or "NEW CALEDONIA" or "NEW ZEALAND").tw.
50	(NICARAGUA or NIGER or NIGERIA or NIUE or "NORFOLK ISLAND" or "NORTHERN MARIANA ISLANDS" or NORWAY or OMAN or PAKISTAN or PALAU or PALESTINE or PANAMA or "PAPUA NEW GUINEA" or PARAGUAY or PERU or PHILIPPINES or PITCAIRN or POLAND or PORTUGAL or PUERTO RICO or QATAR or REUNION or ROMANIA or "RUSSIAN FEDERATION").tw.
51	(RWANDA or "SAINT BARTHÉLEMY" or "SAINT HELENA" or ASCENSION or "TRISTAN DA CUNHA" or "SAINT KITTS" or "SAINT LUCIA" or "SAINT MARTIN" or "SAINT PIERRE" or "SAINT VINCENT" or SAMOA or "SAN MARINO" or "SAO TOME" or "SAUDI ARABIA" or SENEGAL or SERBIA or SEYCHELLES or "SIERRA LEONE" or SINGAPORE or "SINT MAARTEN").tw.
52	(SLOVAKIA or SLOVENIA or "SOLOMON ISLANDS" or SOMALIA or "SOUTH AFRICA" or "SOUTH GEORGIA" or SUDAN or SPAIN or "SRI LANKA" or SUDAN or SURINAME or SVALBARD or SWAZILAND or SWEDEN or SWITZERLAND or "SYRIAN ARAB REPUBLIC" or TAIWAN or TAJIKISTAN or TANZANIA or THAILAND or TIMOR-LESTE or TOGO or TOKELAU or TONGA).tw.
53	(TRINIDAD or TOBAGO or TUNISIA or TURKEY or TURKMENISTAN or "TURKS AND CAICOS ISLANDS" or TUVALU or UGANDA or UKRAINE or "UNITED ARAB EMIRATES" or "UNITED KINGDOM" or "UNITED STATES" or "URUGUAY" or UZBEKISTAN or VANUATU or VENEZUELA or "VIET NAM" or "VIRGIN ISLANDS" or "WALLIS AND FUTUNA").tw.
54	("WESTERN SAHARA" or YEMEN or ZAMBIA or ZIMBABWE or Great Britain or England or Scotland or Wales or America* or Syria or RUSSIA or Vietnam).tw.
55	or/31-54
56	21 AND 30 AND 55

Embase 1947-Present, updated daily via OVID

1	exp dementia/ep [Epidemiology]
2	exp Alzheimer disease/ep [Epidemiology]
3	exp delirium/ep [Epidemiology]
4	exp cognitive defect/ep [Epidemiology]
5	exp Wernicke encephalopathy/ep [Epidemiology]
6	dement*.ti,ab.
7	alzheimer*.ti,ab.
8	(lewy* adj2 bod*).ti,ab.
9	deliri*.ti,ab.
10	(chronic adj2 cerebrovascular).ti,ab.
11	"organic brain disease".ti,ab.
12	"organic brain syndrome".ti,ab.
13	("normal pressure hydrocephalus" and "shunt*").ti,ab.
14	"benign senescent forgetfulness".ti,ab.
15	(cerebr* adj2 deteriorat*).ti,ab.
16	(cerebr* adj2 insufficient*).ti,ab.
17	(pick* adj2 disease).ti,ab.
18	huntington*.ti,ab.
19	binswanger*.ti,ab.
20	korsako*.ti,ab.
21	or/1-20
22	exp incidence/
23	exp prevalence/
24	Epidemiologic Studies/
25	Incidence.ti,ab.
26	Prevalence.ti,ab.
27	distribut*.ti,ab.
28	epidem*.ti,ab.
29	"geographical variation*".ti,ab.
30	or/22-29
31	21 AND 30
32	(Geograph* or cluster* or AFGHANISTAN or "ÅLAND ISLANDS" or ALBANIA or SAMOA or ANDORRA or ANGOLA or ANGUILLA or ANTARCTICA or ANTIGUA or BARBUDA or ARGENTINA or ARMENIA or ARUBA or AUSTRALIA or AZERBAIJAN or BAHAMAS or BANGLADESH or BARBADOS or BELGIUM or BELIZE or BERMUDA or BOLIVIA or BONAIRE or BOSNIA).tw.
33	(HERZEGOVINA or BOTSWANA or "BOUVET ISLAND" or BRAZIL or "BRITISH INDIAN OCEAN TERRITORY" or BRUNEI or DARUSSALAM or BULGARIA or "BURKINA FASO" or BURUNDI or CAMBODIA or CAMEROON or CANADA or CAPE VERDE or "CAYMAN ISLANDS" or "CENTRAL AFRICAN REPUBLIC" or CHAD or CHILE or CHINA or "CHRISTMAS ISLAND").tw.
34	("COCOS ISLANDS" or "KEELING islands" or COLOMBIA or COMOROS or CONGO or "COOK ISLANDS" or "COSTA CÔTE" or CROATIA or CUBA or CURACAO or CYPRUS or CZECH* or DENMARK or DJIBOUTI or DOMINICA or "DOMINICAN REPUBLIC" or ECUADOR or EGYPT or "EL SALVADOR").tw.

35	("EQUATORIAL GUINEA" or ERITREA or ESTONIA or ETHIOPIA or "FALKLAND ISLANDS" or "FAROE ISLANDS" or FIJI or FINLAND or FRANCE or GUIANA or POLYNESIA or "FRENCH SOUTHERN TERRITORIES" or GABON or GAMBIA or GEORGIA or GERMANY or GHANA or GIBRALTAR or GREECE).tw.
36	(GREENLAND or GRENADA or GUADELOUPE or GUAM or GUATEMALA or GUERNSEY or GUINEA or "GUINEA-BISSAU" or HAITI or "HOLY SEE" or HONDURAS or HONG KONG or HUNGARY or ICELAND or INDIA or INDONESIA or IRAN or IRAQ or IRELAND or "ISLE OF MAN" or ISRAEL or ITALY or JAMAICA or JAPAN or JERSEY or JORDAN or KAZAKHSTAN or KENYA).tw.
37	(KIRIBATI or KOREA or KUWAIT or KYRGYZSTAN or LAO or LATVIA or LEBANON or LESOTHO or LIBERIA or LIBYA or LIECHTENSTEIN or LITHUANIA or LUXEMBOURG or MACAO or "MO MAC" or MACEDONIA or MADAGASCAR or MALAWI or MALAYSIA or MALDIVES or MALI or MALTA).tw.
38	("MARSHALL ISLANDS" or MARTINIQUE or MAURITANIA or MAYOTTE or MEXICO or MICRONESIA or MOLDOVA or MONACO or MONGOLIA or MONTSERRAT or MOROCCO or MOZAMBIQUE or MYANMAR or NAMIBIA or NAURU or NEPAL or NETHERLANDS or "NEW CALEDONIA" or "NEW ZEALAND").tw.
39	(NICARAGUA or NIGER or NIGERIA or NIUE or "NORFOLK ISLAND" or "NORTHERN MARIANA ISLANDS" or NORWAY or OMAN or PAKISTAN or PALAU or PALESTINE or PANAMA or "PAPUA NEW GUINEA" or PARAGUAY or PERU or PHILIPPINES or PITCAIRN or POLAND or PORTUGAL or PUERTO RICO or QATAR or REUNION or ROMANIA or "RUSSIAN FEDERATION").tw.
40	(RWANDA or "SAINT BARTHÉLEMY" or "SAINT HELENA" or ASCENSION or "TRISTAN DA CUNHA" or "SAINT KITTS" or "SAINT LUCIA" or "SAINT MARTIN" or "SAINT PIERRE" or "SAINT VINCENT" or SAMOA or "SAN MARINO" or "SAO TOME" or "SAUDI ARABIA" or SENEGAL or SERBIA or SEYCHELLES or "SIERRA LEONE" or SINGAPORE or "SINT MAARTEN").tw.
41	(SLOVAKIA or SLOVENIA or "SOLOMON ISLANDS" or SOMALIA or "SOUTH AFRICA" or "SOUTH GEORGIA" or SUDAN or SPAIN or "SRI LANKA" or SUDAN or SURINAME or SVALBARD or SWAZILAND or SWEDEN or SWITZERLAND or "SYRIAN ARAB REPUBLIC" or TAIWAN or TAJIKISTAN or TANZANIA or THAILAND or TIMOR-LESTE or TOGO or TOKELAU or TONGA).tw.
42	(TRINIDAD or TOBAGO or TUNISIA or TURKEY or TURKMENISTAN or "TURKS AND CAICOS ISLANDS" or TUVALU or UGANDA or UKRAINE or "UNITED ARAB EMIRATES" or "UNITED KINGDOM" or "UNITED STATES" or "URUGUAY" or UZBEKISTAN or VANUATU or VENEZUELA or "VIET NAM" or "VIRGIN ISLANDS" or "WALLIS AND FUTUNA").tw.
43	("WESTERN SAHARA" or YEMEN or ZAMBIA or ZIMBABWE or Great Britain or England or Scotland or Wales or America* or Syria or RUSSIA or Vietnam).tw.
44	OR/32-43
45	31 AND 44

Applied Social Sciences Index & Abstracts (ASSIA) via Proquest

(ti(Dementia) OR ab(Dementia) OR ti("Alzheimer Disease") OR ab("Alzheimer Disease") OR ti(Delirium) OR ab(Delirium) OR ti("Wernicke Encephalopathy") OR ab("Wernicke Encephalopathy") OR ti("Cognitive Disorder*") OR ab("Cognitive Disorder*") OR ti(huntington*) OR ab(huntington*) OR ti(Aphasia) OR ab(Aphasia) OR ti(Creutzfeldt-Jakob) OR ab(Creutzfeldt-Jakob) OR ti(CADASIL) OR ab(CADASIL) OR ti(Lewy Body) OR ab(Lewy Body) OR ti(Pick Disease) OR ab(Pick Disease)) AND (ti(Prevalence) OR ab(Prevalence) OR ti(incidence) OR ab(incidence) OR ti(distribut*) OR ab(distribut*) OR ti(epidem*) OR ab(epidem*)) AND (ti(Geograph*) OR ab(Geograph*) OR ti(cluster*) OR ab(cluster*) OR ti(Afghanistan) OR ab(Afghanistan) OR ti(Albania) OR ab(Albania) OR ti(Algeria) OR ab(Algeria) OR ti(Angola) OR ab(Angola) OR ti(Antarctica) OR ab(Antarctica) OR ti(Argentina) OR ab(Argentina) OR ti(Armenia) OR ab(Armenia) OR ti(Australia) OR ab(Australia) OR ti(Austria) OR ab(Austria) OR ti(Azerbaijan)

OR ab(Azerbaijan) OR ti(Bahamas) OR ab(Bahamas) OR ti(Bahrain) OR ab(Bahrain) OR
ti(Bangladesh) OR ab(Bangladesh) OR ti(Barbados) OR ab(Barbados) OR ti(Belarus) OR ti(Belgium)
OR ab(Belgium) OR ti(Belize) OR ab(Belize) OR ti(Benin) OR ab(Benin) OR ti(Bermuda) OR
ab(Bermuda) OR ti(Bolivia) OR ab(Bolivia) OR ti(Bosnia) OR ab(Bosnia) OR ti(Botswana) OR
ab(Botswana) OR ti(Brazil) OR ab(Brazil) OR ti(Brunei) OR ab(Brunei) OR ti(Bulgaria) OR
ab(Bulgaria) OR ti(Burkina fasso) OR ab(Burkina fasso) OR ti(Burma) OR ab(Burma) OR
ti(Burundi) OR ab(Burundi) OR ti(Cambodia) OR ab(Cambodia) OR ti(Cameroon) OR
ab(Cameroon) OR ti(Canada) OR ab(Canada) OR ti(Central African republic) or ab(Central African
republic) OR ti(Chad) OR ab(Chad) OR ti(Chile) OR ab(Chile) OR ti(China) OR ab(China) OR
ti(Colombia) OR ab(Colombia) OR ti(Congo) OR ab(Congo) OR ti(Costa Rica) OR ab(Costa Rica)
Or ti(Cote divoire) OR ab(Cote divoire) OR ti(Croatia) OR ab(Croatia) OR ti(Cuba) OR ab(Cuba) OR
ti(Cyprus) OR ab(Cyprus) OR ti(Czech republic) OR ab(Czech republic) OR ti(Denmark) OR
ab(Denmark) OR ti(Dominica) OR ab(Dominica) OR ti(Dominican republic) OR ab(Dominican
republic) OR ti(Ecuador) OR ab(Ecuador) OR ti(Egypt) OR ab(Egypt) OR ti(El Salvador) OR ab(El
Salvador) OR ti(England) OR ab(England) OR ti(Equatorial guinea) OR ab(Equatorial guinea) OR
ti(Eritrea) OR ab(Eritrea) OR ti(Estonia) OR ab(Estonia) OR ti(Ethiopia) OR ab(Ethiopia) OR
ti(Finland) OR ab(Finland) OR ti(France) OR ab(France) OR ti(Georgia) OR ab(Georgia) OR
ti(Germany) OR ab(Germany) OR ti(Ghana) OR ab(Ghana) OR ti(Greece) OR ab(Greece) OR
ti(Greenland) OR ab(Greenland) OR ti(Guinea) OR ab(Guinea) OR ti(Hong kong) OR ab(Hong kong)
OR ti(Hungary) OR ab(Hungary) OR ti(Iceland) OR ab(Iceland) OR ti(India) OR ab(India) OR
ti(Indonesia) OR ab(Indonesia) OR ti(Iran) OR ab(Iran) OR ti(Iraq) OR ab(Iraq) OR ti(Ireland) OR
ab(Ireland) OR ti(Israel) OR ab(Israel) OR ti(Italy) OR ab(Italy) OR ti(Jamaica) OR ab(Jamaica) OR
ti(Japan) OR ab(Japan) OR ti(Jordan) OR ab(Jordan) OR ti(Kazakhstan) OR ab(Kazakhstan) OR
ti(Kenya) OR ab(Kenya) OR ti(Korea) OR ab(Korea) OR ti(Kuwait) OR ab(Kuwait) OR
ti(Kyrgyzstan) OR ab(Kyrgyzstan) OR ti(Laos) OR ab(Laos) OR ti(Latvia) OR ab(Latvia) OR
ti(Lebanon) OR ab(Lebanon) OR ti(Lesotho) OR ab(Lesotho) OR ti(Libya) OR ab(Libya) OR
ti(Lithuania) OR ab(Lithuania) OR ti(Malawi) OR ab(Malawi) OR ti(Malaysia) OR ab(Malaysia) OR
ti(Mali) OR ab(Mali) OR ti(Mexico) OR ab(Mexico) OR ti(Mongolia) OR ab(Mongolia) OR
ti(Morocco) OR ab(Morocco) OR ti(Mozambique) OR ab(Mozambique) OR ti(Namibia) OR
ab(Namibia) OR ti(Nepal) OR ab(Nepal) OR ti(Netherlands) OR ab(Netherlands) OR ti(New
Zealand) OR ab(New Zealand) OR ti(Nicaragua) OR ab(Nicaragua) OR ti(Niger) OR ab(Niger) OR
ti(Nigeria) OR ab(Nigeria) OR ti(Norway) OR ab(Norway) OR ti(Pakistan) OR ab(Pakistan) OR
ti(Papua new Guinea) OR ab(Papua new Guinea) OR ti(Paraguay) OR ab(Paraguay) OR ti(Peru) OR
ab(Peru) OR ti(Philippines) OR ab(Philippines) OR ti(Poland) OR ab(Poland) OR ti(Portugal) OR
ab(Portugal) OR ti(Puerto Rico) OR ab(Puerto Rico) OR ti(Romania) OR ab(Romania) OR ti(Russia)
OR ab(Russia) OR ti(Rwanda) OR ab(Rwanda) OR ti(Saudi Arabia) OR ab(Saudi Arabia) OR
ti(Scotland) OR ab(Scotland) Or ti(Senegal) OR ab(Senegal) OR ti(Serbia) OR ab(Serbia) OR
ti(Sierra Leone) OR ab(Sierra Leone) OR ti(Singapore) OR ab(Singapore) OR ti(Slovakia) OR
ab(Slovakia) OR ti(Slovenia) OR ab(Slovenia) OR ti(Somalia) OR ab(Somalia) OR ti(South Africa)
OR ab(South Africa) OR ti(Spain) Or ab(Spain) OR ti(Sri lanka) OR ab(Sri lanka) OR ti(Sudan) OR
ab(Sudan) OR ti(Swaziland) OR ab(Swaziland) OR ti(Sweden) OR ab(Sweden) OR ti(Switzerland)
OR ab(Switzerland) OR ti(Syria) OR ab(Syria) OR ti(Taiwan) OR ab(Taiwan) OR ti(Tajikistan) OR
ab(Tajikistan) OR ti(Tanzania) OR ab(Tanzania) OR ti(Thailand) OR ab(Thailand) OR ti(Togo) OR
ab(Togo) OR ti(Trinidad) OR ab(Trinidad) OR ti(Tunisia) OR ab(Tunisia) OR ti(Tunisia) OR
ab(Tunisia) OR ti(Turkey) OR ab(Turkey) OR ti(Turkmenistan) OR ab(Turkmenistan) OR ti(Uganda)
OR ab(Uganda) OR ti(Ukraine) OR ab(Ukraine) OR ti(United Arab Emirates) OR ab(United Arab
Emirates) OR ti(United kingdom) OR ab(United kingdom) OR ti(United states) OR ab(United states)
OR ti(Uruguay) OR ab(Uruguay) OR ti(Uzbekistan) OR ab(Uzbekistan) OR ti(Venezuela) OR
ab(Venezuela) OR ti(Vietnam) OR ab(Vietnam) OR ti(Wales) OR ab(Wales) OR ti(Zambia) OR
ab(Zambia) OR ti(Zimbabwe) OR ab(Zimbabwe))

Appendix 2

Eligibility criteria

Included:

- Cross-sectional or longitudinal studies of any duration comparing primary dementia rates (prevalence or incidence) between two or more different geographical sites, at any geographical scale but including a comparison of rural and urban areas

Excluded:

- Studies were no comparison between two or more different geographical sites or if the comparison between groups did not use the same methodology
- Studies where dementia was a secondary feature of a chronic disease (alcohol or traumatic brain injury), Parkinson's disease, Huntington's disease and Creutzfeldt-Jakob disease.
- Studies purely on young onset dementia (<60 years old) or mild cognitive impairment. We did not restrict to English language.

Appendix 3: Table of country classification based on income and different years/classification

	Wellcome research classification of LIC and LMIC countries based on The Organisation for Economic Co- operation and Development (OECD)*	World Bank Definition 2018 Countries	World Bank Definition 2022-2023 Countries
High Income Country (HIC)	USA, UK, Canada, Australia, Switzerland, Spain, Portugal, Germany, Romania, South Korea, Taiwan,	USA, UK, Canada, Australia, Switzerland, Spain, Portugal, Germany, South Korea, Taiwan	USA, UK, Canada, Australia, Switzerland, Spain, Portugal, Germany, Romania, South Korea, Taiwan
Upper Middle-Income Country (UMIC)		Romania, Peru, Mexico, Brazil, Costa Rica, China, Malaysia, Iran	Peru, Mexico, Brazil, Costa Rica, China, Malaysia
Lower Middle-Income Country (LMIC)	China, Central African Republic (CAR), Republic of Congo (ROC),	Republic of Congo (ROC)	India, Iran, Republic of Congo (ROC), Egypt
Low Income Country (LIC)	Costa Rica, Mexico, India, Peru, Brazil, Iran, Egypt, Malaysia,	India, Central African Republic (CAR), Egypt	Central African Republic (CAR)

¹ *We used this definition for this systematic review. For the meta-analysis, we compared all in the High Income Countries (HIC) as these were largely consistent across the years with all other Income countries (Upper Middle-Income, Lower Middle-Income and Low Income Countries) and referred to this group as Low-to-Middle Income Countries (L-MIC)

Appendix 4: Quality Assessment – Adapted ROBANS

Risk of Bias Item	Grading	Rationale/Examples
Selection bias	Low	Sampling frame clearly described with reasonable inclusion/exclusion criteria provided, ideally allowing for inclusion of all participants within the geographical area being studied
	High	Exclusions based on availability of data (e.g.); studies only considering certain postcodes; exclusion criteria likely to reduce the representativeness of a possible area (e.g. only looking at comparing ethnic groups in one area and not taking into consideration place of birth); only including population if already part of another study; only including population if clinical or in certain clinical area (e.g. comparing General practitioner localities, only comparing catchment areas for hospitals, only comparing geographical areas with large number of nursing homes).
	Unclear	Sampling frame unclear, criteria for inclusion/exclusion not provided or explained
Confounding variables	Low	Multivariate model accounting for likely possible confounding variables
	High	No consideration of confounding variables; univariate analyses only (e.g., areas with non-varied socioeconomic status that would influence risk factor for dementia separately to geographical factors; age; vascular risk factors; large number of GP surgeries or hospitals in vicinity; large number of nursing homes in vicinity).

Measurement of exposure <i>(performance bias)</i>	Unclear	Methods for analysis not clearly described or reported
	Low	Clearly described method on how data were collected and extracted
		Best practice includes description of who performed data extraction, case definitions/descriptions of eligible conditions (e.g. at point during the person's life was the geography measured; how it was measured; who collected the information; what detail was collected – not just 'town').
	High	Missing data on key predictor variables
Incomplete outcome data <i>(Attrition bias)</i>	Unclear	Methods for assessing predictor variables not clearly described
	Low	Outcomes assessed for all included participants (including how many included had cognition assessed/obtained and how many it was missing for) and what method (e.g., MMSE, electronic data with coding for dementia; other cognitive assessments; more than one cognitive assessment)
	High	Missing outcome assessments (e.g. no data on how many did not have cognition assessed/recorded; no data on how cognition was assessed/obtained)
	Unclear	Outcome assessment reported as percentages without absolute values being presented, preventing assessment of completeness of outcome reporting
Selective outcome reporting <i>(reporting bias)</i>	Low	Reporting as per published protocol
	High	Evidence that reporting deviates from publically accessible protocol
	Unclear	No protocol publically available

Appendix 5: Supplementary Table 1 – Additional information on methodologies and range of diagnostic tools used

Author	Methodology
Abner 2016* ¹	Dementia in Electronic health records.
Astell_Burt 2018* ²	Data linkage of records for cholinesterase inhibitor prescriptions.
Chammartin 2016 ³	Dementia from death certificates ICD codes G30-G309
	Ontario Population Health and Environment Cohort (ONPHEC) Incidence study. Dementia from hospital admission/physician claims/prescription (ICD codes).
Chen 2017 ^{4*}	
Chen 2012 ⁵ Anhui study (a)*	Survey/interviews. Dementia defined using the GMS and the Automated Geriatric Examination for Computer Assisted Taxonomy (AGECAT).
Chen 2012 ⁵ Four provinces study (b)*	Survey/interviews. Dementia defined using the GMS and the Automated Geriatric Examination for Computer Assisted Taxonomy (AGECAT).
	Two cross sectional incidence surveys, (1994-1995,1997-1998). Screened with WHO protocol for dementia and confirmed by clinical assessment.
Contador 2015 ⁶	Childhood/adulthood residence and dementia risk investigated.
Cornutiu 2010 ⁷	Incidence study, data from annual hospitalizations for AD. NINCDS_ADRDA criteria. The study population was unclear as possibly it included 600 participants but also documented 246 in 2002.
Deng 2018* ⁸	Questionnaire/interview & clinical assessment. MMSE, IADL, GDS.
Fronza da Silva 2013 ⁹	Study investigated pesticide use in rural and urban areas and dementia. Questionnaire by clinician. Carer completed if participant could not.
Ganesh 2010 ¹⁰	Thesis. Medicare Current Beneficiary Survey (MCBS) ICD-9-CM codes: claims or self-report for AD.
Goncalves-Pereira 2017* ¹¹	Cross-sectional survey. Dementia using 10/66 Dementia Research Group or DSM IV criteria. 10/66 data used for meta-analysis.

Goodman 2017* ¹²	Prevalence study from Centers for Medicare & Medicaid (CMS) claims data (ICD-9-CM)
Guerchet (a) 2013* ¹³	Interview/survey study in EPIDEMCA population: Community Screening Interview for Dementia (CSI-D), the GMS-AGECAT and the CERAD's 10-word list. DSM IV criteria used. 10/66 data not complete. Additional information from author allowing inclusion in meta-analysis.
Guerchet (b) 2013* ¹³	Interview/survey study in EPIDEMCA population: Community Screening Interview for Dementia (CSI-D), the GMS-AGECAT and the CERAD's 10-word list. DSM IV criteria used. 10/66 data not complete. Additional information from author allowing inclusion in meta-analysis.
Hendrie 2018* ¹⁴	Indianapolis-Ibadan project (IIDP) incidence study. Only Indianapolis people included in this study. Two cohorts (1992 and 2001). Investigated childhood/adulthood residence and dementia risk. Clinical evaluation of dementia using neuropsychological tests, standardized neurologic and physical examination and functional status review and structured interview from informant. Medical records also reviewed.
Huang 2014 ¹⁵	Newly diagnosed patients with diabetes (n= 71,433) and matched controls (n= 71,311). Electronic health records using Taiwan's National Health Insurance Research Database (AD claims (ICD-9-CM code: 331.0). We used 'Metropolis' in their definitions to extract the data for Urban location.
Jia 2014 ^{16*}	China Cognition and Ageing Study (China Coast) cohort: cross-sectional survey/clinical assessment. Clinical Dementia Rating (CDR) score
Khedr 2015 ^{17*}	Cross-sectional community survey & clinical assessment. DSM-IV diagnostic criteria.
Kim 2011* ¹⁸	Population survey (MMSE-KC) then clinical evaluation.
Koller 2010 ^{19*}	Diagnosis in Clinical records

Nadel 2014 ²⁰	Survey. MMSE & structured interview. ‘Severe cognitive impairment’ appeared to be synonymous with dementia so this information was extracted for the dementia outcome.
Nunes 2010 ^{*21}	Survey & clinical assessment.
Pilleron 2015 ²² (a)*	Interview/survey study in EPIDEMCA population: DSM IV criteria. Cognitive testing using Community Screening Interview for Dementia (CSI-D). Informant (relative) to assess daily activities. Further subsequent clinical assessment. Further psychometric tests (Free and Cued Selective Reminding Test, Zazzo’s cancellation task and Issac’s Set Test of Verbal Fluency).
Pilleron 2015 ²² (b)*	Interview/survey study in EPIDEMCA population: DSM IV criteria. Cognitive testing using Community Screening Interview for Dementia (CSI-D). Informant (relative) to assess daily activities. Further subsequent clinical assessment. Further psychometric tests (Free and Cued Selective Reminding Test, Zazzo’s cancellation task and Issac’s Set Test of Verbal Fluency).
Poddar 2011 ^{23*}	Cross-sectional. Two phase- phase 1 interview/survey and phase 2 - Hindi Mental State Examination
Prince 2012 ^{24*} (a)	The countries that had both urban and rural totals were Mexico, Peru and China. This is the data for the Peru study. 10/66 Dementia Research Group population-based studies; interview. Community Screening Instrument for Dementia (CSI‘D’) COGSCORE. 2003-2007 one-phase population based surveys and then 2007-2010 for the incidence wave and follow up. Both 10/66 and DSM-IV criteria were used for establishing cases of dementia.

Prince 2012 ^{24*} (b)	The countries that had both urban and rural totals were Mexico, Peru and China. This is the data for the Mexico study. 10/66 Dementia Research Group population-based studies; interview. Community Screening Instrument for Dementia (CSI'D') COGSCORE. 2003-2007 one-phase population-based surveys and then 2007-2010 for the incidence wave and follow up. Both 10/66 and DSM-IV criteria were used for establishing cases of dementia.
Prince 2012 ^{24*} (c)	The countries that had both urban and rural totals were Mexico, Peru and China. This is the data for the China study. 10/66 Dementia Research Group population-based studies; interview. Community Screening Instrument for Dementia (CSI'D') COGSCORE. 2003-2007 one-phase population-based surveys and then 2007-2010 for the incidence wave and follow up. Both 10/66 and DSM-IV criteria were used for establishing cases of dementia.
Raina 2014 ^{25*}	No cohort was used. Two-phase process (screening and clinical phase). (Hindi Mental State Examination (HMSE) and modified version for tribal group, clinical assessment). Study also investigated tribal and migrant populations but excluded for the meta-analysis and only rural and urban included.
Rodríguez-Agudelo 2011 ^{26*}	Part of the 10/66 Research Group studies. Investigating neuropsychiatric symptoms in those with and without dementia. Two areas selected. Initial eligibility interviews at home followed by further interviews at health centres. Group 10/66 detection algorithm used for dementia diagnosis.
Samba 2015 ^{27*}	EPIDEMCA (Epidemiology of Dementia in Central Africa) - FU (Follow-up) study cohort. Longitudinal population-based incidence study (966 participants at baseline without dementia). Participants traced with

Psychometrical tests and health and vascular events investigated since the last visit. Verbal autopsies performed when older person was deceased and a reliable information available. DSM-IV and NINCDS-ADRDA criteria used for dementia/Alzheimer disease diagnoses.

Sharifi 2016^{28*}

Two-phase screening cross-sectional study with multistage cluster random sampling national survey (The National Elderly Health Survey, NEHS). Physical examination and interview. Dementia diagnosis in two phases: Brief Cognitive Assessment Tool (BCAT) from Iran and diagnosis of dementia by GPs based on DSM IV criteria. Those with severe depression were removed to ensure no 'pseudo-dementia'.

Taylor 2017²⁹

The study population was unclear as this study investigated AD mortality for different years, but it was unclear what the overall population (denominator) was for those years. State-level and county-level death certificate data from the National Vital Statistics System were analysed analysing ICD 10 codes.

Tengku Aizan
2010^{30*}

The mental Health and Quality of Life of Older Malaysians cohort. Multi-stage sampling door-to-door survey with face-to-face interviews. Geriatric Mental State (GMS) B3 version assessed mental health status and entered into AGE CAT software programme. AGE CAT-GMS diagnostic algorithm was used for dementia cases. Rural and urban total dementia cases were reported separately. For the purpose of this systematic review, we totalled both of these for the total dementia *n* in this table.

Tola-Arribas 2013^{31*}

DEMINVALL cohort (cross-sectional, two-phase, door-to-door, population-based study from social security healthcare holders' registry). Phase 1: Screening tool (Spanish version 7-minute Screen Neurocognitive Battery (7MS) 1, or Spanish version IQCODE, or Spanish version Kawas Dementia Questionnaire if deceased). Phase 2: clinical assessment

[including Spanish version of Cambridge Examination of Mental Disorders of the Elderly and Clinical Dementia Rating (CDR)]

Tornau 2015 ³²	The study population was unclear as the database was stated as having >20,000 patients. Billing data of all statutory health insurances (90% of population) from an urban and a rural region in Germany. ICD 10-GM codes
Tripathi 2012 ^{33*}	Case control study in outpatient tertiary clinic (150 consecutive patients attending cognitive disorders clinic and 150 age and sex matched control subjects attending the outpatient as other patients' relatives). Folstein's Mini Mental Status Examination and clinical assessment (neurologist and neuropsychologist) including AIIMS neuropsychological battery (validated in Indian population) and MRI brain.
Weden 2018 ³⁴ (a)*	Two cohorts within the one paper – 2000 and 2010. This is the 2000 cohort. Health and Retirement Study (HRS) multi-panel cohort panel survey linked to Census assessments of urbanicity. Cognitive functioning measured using validated methodology or if self-reporting using 27-point modified Telephone Interview for Cognitive Status. A proxy respondent used for those unable/unwilling to answer.
Weden 2018 ³⁴ (b)*	Two cohorts within the one paper – 2000 and 2010. This is the 2010 cohort. Health and Retirement Study (HRS) multi-panel cohort panel survey linked to Census assessments of urbanicity. Cognitive functioning measured using validated methodology or if self-reporting using 27-point modified Telephone Interview for Cognitive Status. A proxy respondent used for those unable/unwilling to answer.
Wu 2017 ^{35*}	Baseline interviews were collected between 2008-2011. Urban was defined as urban city and town and also conurbation. For the purpose of this systematic review, urban city and town and conurbation were totalled to

give the Urban $n = 5951$. Cohort study, Cognitive Function and Ageing Study II (CFAS II). Secondary data analysis CFAS II population-based epidemiologic study - baseline interviews, primary care registration. Geriatric Mental Status, MMSE, and Automatic Geriatric Examination for Computer-Assisted Taxonomy organicity level ≥ 3 .

Yin 2016³⁶

Not a cohort. Mortality study using ICD codes from China Mortality Surveillance System Disease Surveillance Point (DSP). The total study participants were unclear.

Yuan 2016^{37*}

Baseline participation was 1997 with follow up in 1999 and then 2000-2002. Cohort study [subsample of the 1997 study Dementia subtypes in China: Prevalence in Beijing (northeast), Xi'an (northwest), Shanghai (southeast), and Chengdu, (southwest)]. Incidence study (three phase). Phase 1 (dementia screening), door-to-door screening survey using the Chinese version of MMSE (C-MMSE) and Chinese version of ADL (CADL). Phase II (clinical diagnostic workup) clinical interview/examination, then re-screening, then MDT discussion. The total study population n in this table were the total traced at the second follow-up period (2000-2002). The Urban n listed in this table comprised the total traced at the second follow-up period (2000-2002).

Zhao 2010³⁸

Three phases: screening, clinical diagnostic ascertainment (clinical assessment and battery of neuropsychological tests - including ADL, POD, FOM, VFT, BD, DS, HAMD, Hachinski Ischemia Scale, C-MMSE, GDS), clinical confirmation (repeated screening and diagnostic examinations at 6 months in phase II participants).

AD = Alzheimer's Disease; ADRD = Alzheimer's Related Dementia; N/A = not available (unable to extract). Grey = not in meta-analysis

Supplementary Table 1 shows the range of diagnostic tools used. Few identified dementia subtype. Most papers used a broad dementia label as their outcome, but some did investigate the specific subtype, such as Alzheimer or vascular dementia. The most used diagnostic tools for dementia ascertainment were surveys/interviews, clinical assessment and ICD codes.

However, for some they separated the groups into those with dementia, those with cognitive impairment/MCI and those without either. Other papers did not specifically separate MCI from dementia; therefore, some within the 'no dementia' group may have MCI. For the purpose of this study, we used those with probable dementia/dementia as the outcome and in one paper 'severe cognitive impairment' appeared synonymous with dementia so this was used.

References

1. Abner EL, Jicha GA, Christian WJ, Schreurs BG. Rural-urban differences in Alzheimer's disease and related disorders diagnostic prevalence in Kentucky and West Virginia. *Journal of Rural Health* 2016;**32**(3):314-320.
2. Astell-Burt T, Feng X. Is the risk of developing Alzheimer's disease really higher in rural areas? A multilevel longitudinal study of 261,669 Australians aged 45 years and older tracked over 11 years. *Health and Place* 2018;**54**:132-137.
3. Chammartin F, Probst-Hensch N, Utzinger J, Vounatsou P. Mortality atlas of the main causes of death in Switzerland, 2008-2012. *Swiss Medical Weekly* 2016;**146**:w14280.
4. Chen H, Kwong JC, Copes R, et al. Exposure to ambient air pollution and the incidence of dementia: A population-based cohort study. *Environment International* 2017;**108**:271-277.
5. Chen R, Ma Y, Wilson K, et al. A multicentre community-based study of dementia cases and subcases in older people in China--the GMS-AGECAT prevalence and socio-economic correlates. *International Journal of Geriatric Psychiatry* 2012;**27**(7):692-702.
6. Contador I, Bermejo-Pareja F, Puertas-Martin V, Benito-Leon J. Childhood and Adulthood Rural Residence Increases the Risk of Dementia: NEDICES Study. *Current Alzheimer Research* 2015;**12**(4):350-7.
7. Cornutiu G. The incidence and prevalence of Alzheimer's disease. *Neurodegenerative Diseases* 2010;**8**(1-2):9-14.
8. Deng J, Cao C, Jiang Y, et al. Prevalence and effect factors of dementia among the community elderly in Chongqing, China. *Psychogeriatrics* 2018;**18**(5):412-420.
9. Silva EFd, Paniz VMV, Laste G, Torres ILdS. Prevalência de morbidades e sintomas em idosos: um estudo comparativo entre zonas rural e urbana. *Ciência & Saúde Coletiva* 2013;**18**(4):1029-1040.
10. Ganesh C. The diffusion of prescription drugs for Alzheimer's disease among medicare beneficiaries [Ph.D.]. The Pennsylvania State University, 2010.

11. Goncalves-Pereira M, Cardoso A, Verdelho A, et al. The prevalence of dementia in a Portuguese community sample: a 10/66 Dementia Research Group study. *BMC Geriatrics* 2017;**17**(1):261.
12. Goodman RA, Lochner KA, Thambisetty M, et al. Prevalence of dementia subtypes in United States Medicare fee-for-service beneficiaries, 2011-2013. *Alzheimer's & Dementia* 2017;**13**(1):28-37.
13. Guerchet M, Ndamba-Bandzouzi B, Mbelesso P, et al. Comparison of rural and urban dementia prevalences in two countries of central africa: The epidemca study. *Alzheimer's and Dementia* 2013;**1**:P688.
14. Hendrie HC, Smith-Gamble V, Lane KA, et al. The Association of Early Life Factors and Declining Incidence Rates of Dementia in an Elderly Population of African Americans. *Journals of Gerontology Series B: Psychological Sciences & Social Sciences* 2018;**73**:S82-S89.
15. Huang C, Chung C, Leu H, et al. Diabetes mellitus and the risk of Alzheimer's disease: a nationwide population-based study. *PLoS ONE* 2014;**9**(1).
16. Jia J, Wang F, Wei C, et al. The prevalence of dementia in urban and rural areas of China. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2014;**10**(1):1-9.
17. Khedr E, Fawi G, Abbas MA, et al. Prevalence of mild cognitive impairment and dementia among the elderly population of Qena Governorate, Upper Egypt: a community-based study. *Journal of Alzheimer's Disease* 2015;**45**(1):117-26.
18. Kim K, Park J, Kim M, et al. A nationwide survey on the prevalence of dementia and mild cognitive impairment in South Korea. *JAD, Journal of Alzheimer's Disease* 2011;**23**(2):281-291.
19. Koller D, Eisele M, Kaduszkiewicz H, et al. Ambulatory health services utilization in patients with dementia - is there an urban-rural difference? *International Journal of Health Geographics* 2010;**9**(59).
20. Nadel JL, Ulate D. Incidence and risk factors for cognitive impairment in rural elderly populations in Costa Rica^ien

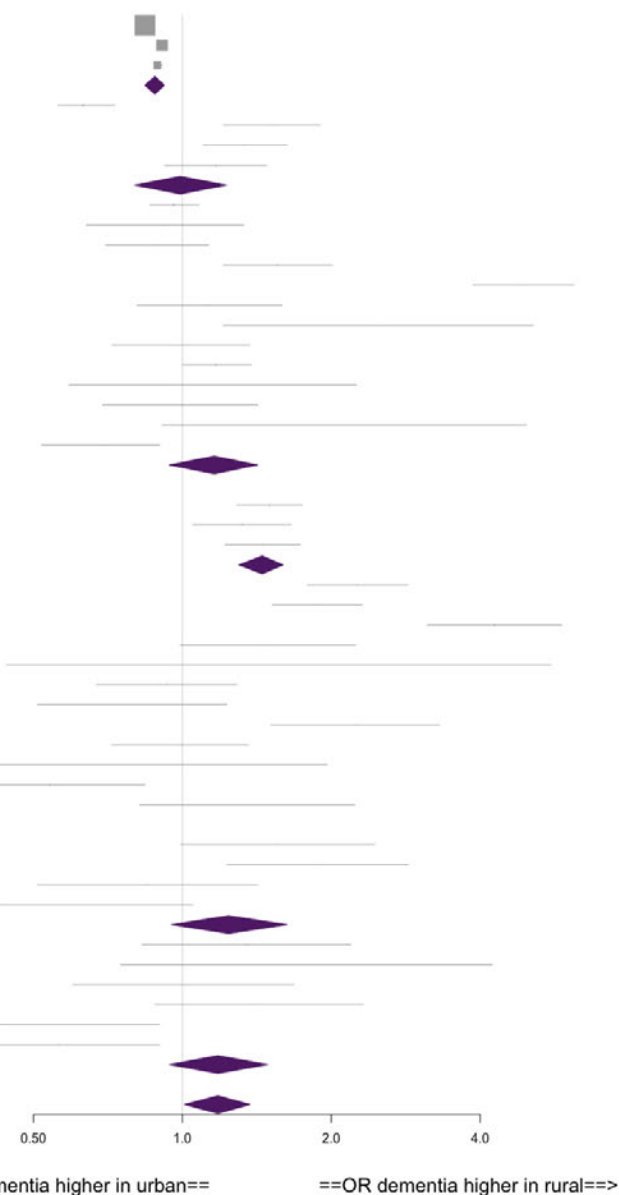
Incidencia y factores de riesgo para la discapacidad cognitiva en poblaciones rurales de tercera edad en Costa Rica^ies. *Rev. biol. trop* 2014;**62**(3):869-876.

21. Nunes B, Silva RD, Cruz VT, et al. Prevalence and pattern of cognitive impairment in rural and urban populations from Northern Portugal. *BMC Neurology Vol 10 2010, ArtID 42* 2010;**10**.
22. Pilleron S, Clement JP, Ndamba-Bandzouzi B, et al. Is dependent personality disorder associated with mild cognitive impairment and dementia in Central Africa? A result from the EPIDEMCA programme. *International Psychogeriatrics* 2015;**27**(2):279-288.
23. Poddar K, Kant S, Singh A, Singh TB. An epidemiological study of dementia among the habitants of Eastern Uttar Pradesh, India. *Annals of the Indian Academy of Neurology* 2011;**14**(3):164-168.
24. Prince M, Acosta D, Ferri CP, et al. Dementia incidence and mortality in middle-income countries, and associations with indicators of cognitive reserve: A 10/66 Dementia Research Group population-based cohort study. *The Lancet* 2012;**380**(9836):50-58.
25. Raina SK, Raina S, Chander V, et al. Is dementia differentially distributed? A study on the prevalence of dementia in migrant, urban, rural, and tribal elderly population of Himalayan region in Northern India. *North American Journal of Medical Sciences* 2014;**6**(4):172-177.
26. Rodriguez-Agudelo Y, Solis-Vivanco R, Acosta-Castillo I, et al. Neuropsychiatric symptoms in older adulte with and without dementia in urban and rural regions. Results of the 10/66 Dementia Research Group in Mexico. [Spanish]

Sintomas neuropsiquiatricos en adultos mayores con y sin demencia de regiones urbana y rural. Resultados del Grupo de Investigacion en Demencia 10/66 en Mexico. *Revista de Investigacion Clinica* 2011;**63**(4):382-390.

27. Samba H, Guerchet M, Bandzouzi BN, et al. Incidence of dementia among older adults in central africa: First results from the republic of congo in the EPIDEMCA-FU study. *Alzheimer's and Dementia* 2015;**1**:P221-P222.
28. Sharifi F, Fakhrzadeh H, Varmaghani M, et al. Prevalence of Dementia and Associated Factors among Older Adults in Iran: National Elderly Health Survey (NEHS). *Archives of Iranian Medicine* 2016;**19**(12):838-844.
29. Taylor CA, Greenlund SF, McGuire LC, Lu H, Croft JB. Deaths from Alzheimer's Disease - United States, 1999-2014. *Mmwr* 2017;**Morbidity and mortality weekly report**. **66**(20):521-526.
30. Tengku Aizan H, Krishnaswamy S, Siti Suhailah A, Yadollah Abolfathi M. Sociodemographic risk factors and correlates of dementia in older Malaysians. *Dementia and Geriatric Cognitive Disorders* 2010;**30**(6):533-539.
31. Tola-Arribas MA, Yugueros MI, Garea MJ, et al. Prevalence of dementia and subtypes in Valladolid, northwestern Spain: The DEMINVAL Study. *PLoS ONE* 2013;**8**(10).
32. Tornau F, Bruchmann G, Juckel G. Prevalence of psychiatric disorders in the elderly in rural and urban areas in Germany. *International Psychogeriatrics* 2015;**27**:S167.
33. Tripathi M, Vibha D, Gupta P, et al. Risk factors of dementia in North India: A case-control study. *Aging & Mental Health* 2012;**16**(2):228-235.
34. Weden MM, Shih RA, Kabeto MU, Langa KM. Secular trends in dementia and cognitive impairment of U.S. rural and urban older adults. *American Journal of Preventive Medicine* 2018;**54**(2):164-172.
35. Wu YT, Prina AM, Jones A, et al. The Built Environment and Cognitive Disorders: Results From the Cognitive Function and Ageing Study II. *American Journal of Preventive Medicine* 2017;**53**(1):25-32.
36. Yin P, Feng X, Astell-Burt T, et al. Temporal trends and geographic variations in dementia mortality in China between 2006 and 2012: Multilevel evidence from a nationally representative sample. *Alzheimer Disease and Associated Disorders* 2016;**30**(4):348-353.
37. Yuan J, Zhang Z, Wen H, et al. Incidence of dementia and subtypes: A cohort study in four regions in China. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2016;**12**(3):262-271.
38. Zhao Q, Zhou B, Ding D, Guo Q, Hong Z. Prevalence, mortality, and predictive factors on survival of dementia in Shanghai, China. *Alzheimer Disease and Associated Disorders* 2010;**24**(2):151-158.

Study	Country	Total N	Dementia	OR (95% CI)
Goodman 2017	USA	21,624,228	3,110,654	0.84 (0.84, 0.85)
Chen 2017	Canada	2,066,639	257,816	0.91 (0.90, 0.92)
Abner 2016	USA	1,209,976	82,278	0.89 (0.88, 0.91)
High Income Countries: N>1,000,000				0.88 (0.84, 0.92)
Astell-Burt 2018	Australia	261,669	3,046	0.63 (0.56, 0.73)
Jean 1996	Canada	131,667	219	1.52 (1.21, 1.90)
Weden 2018a	USA	16,386	958	1.34 (1.10, 1.63)
Weden 2018b	USA	16,311	727	1.17 (0.92, 1.48)
High Income Countries: N>10,000				0.99 (0.80, 1.23)
Koller 2010	Germany	9,216	1,846	0.96 (0.86, 1.08)
Wu 2017	England	7,505	328	0.92 (0.64, 1.33)
Matthews 2005	UK	3,557	217	0.89 (0.70, 1.13)
Hendrie 2018	USA	3,276	285	1.56 (1.21, 2.01)
Arsalantas 2009	Turkey	3,100	262	4.89 (3.87, 6.19)
Lin 1998	Taiwan	2,915	108	1.13 (0.81, 1.59)
Ogunniyi 2000b	USA	2,212	65	2.49 (1.21, 5.12)
Tola-Arribas 2013	Spain	2,170	184	0.99 (0.72, 1.37)
Kim 2011	South Korea	1,673	351	1.17 (1.00, 1.38)
Yip 1997	Taiwan	1,443	29	1.16 (0.59, 2.25)
Goncalves-Pereira 2017	Portugal	1,405	129	0.99 (0.69, 1.42)
Nunes 2010	Portugal	1,146	31	2.12 (0.91, 4.96)
Azzimondi 1998	Italy	727	196	0.68 (0.52, 0.90)
All High Income Countries				1.16 (0.94, 1.42)
Zhang 2006	China	34,807	1,027	1.50 (1.29, 1.75)
Yuan 2016	China	12,881	311	1.32 (1.05, 1.66)
Jia 2014	China	10,276	528	1.45 (1.22, 1.73)
Low-to-Middle Income Countries: N>10,000				1.45 (1.30, 1.60)
Chen 2012b	China	3,327	341	2.26 (1.79, 2.86)
Tengku Aizan 2010	Malaysia	2,980	418	1.87 (1.52, 2.31)
Chen 2012a	China	2,917	210	4.27 (3.13, 5.84)
Poddar 2011	India	2,890	146	1.49 (0.99, 2.24)
Ogunniyi 2000a	Nigeria	2,494	28	1.57 (0.44, 5.56)
Prince 2012c	China	2,162	161	0.93 (0.67, 1.29)
Rodriguez 2008c	China	2,162	59	0.79 (0.51, 1.23)
Prince 2012b	Mexico	2,003	121	2.24 (1.51, 3.31)
Rodríguez-Agudelo 2011	Mexico	2,003	171	0.99 (0.72, 1.36)
Rodriguez 2008d	India	1,985	17	0.88 (0.39, 1.96)
Rodriguez 2008b	Mexico	2,002	62	0.54 (0.35, 0.84)
Prince 2012a	Peru	1,933	69	1.36 (0.82, 2.23)
Rodriguez 2008a	Peru	1,931	43	0.12 (0.04, 0.39)
Deng 2018	China	1,781	195	1.56 (0.99, 2.45)
Sharifi 2016	Iran	1,257	99	1.88 (1.23, 2.86)
Guerchet 2013b	ROC	1,029	63	0.85 (0.51, 1.42)
Raina 2014	India	1,000	23	0.43 (0.18, 1.05)
Low-to-Middle Income Countries: N>1000				1.24 (0.95, 1.63)
Guerchet 2013a	CAR	973	72	1.35 (0.83, 2.19)
Samba 2015	ROC	966	23	1.78 (0.75, 4.23)
Pilleron 2015b	ROC	912	135	1.01 (0.60, 1.68)
Pilleron 2015a	CAR	860	135	1.43 (0.88, 2.32)
Khedr 2015	Egypt	691	35	0.44 (0.22, 0.90)
Tripathi 2012	India	300	150	0.56 (0.35, 0.90)
All Low-to-Middle Income Countries				1.18 (0.94, 1.49)
All Studies				1.18 (1.01, 1.37)



Appendix 6: Supplementary Figure – Forest plot of all studies in current systematic review and previous 2012 systematic review

Appendix 7

Supplementary Figure 3: Risk of Bias assessment by study

	Risk of bias domains				
	D1	D2	D3	D4	D5
Abner 2016	High	Low	Unclear	High	Unclear
Astell-Burt 2018	High	Low	Low	Low	Unclear
Chammartin 2016	High	High	Low	High	Unclear
Chen 2017	Low	Low	High	Low	Unclear
Chen 2012	Unclear	Low	Unclear	High	Unclear
Contador 2015	Low	Low	Low	Unclear	Unclear
Cornutiu 2011	High	High	High	Unclear	Unclear
Deng 2018	Low	Low	Unclear	Unclear	Unclear
Fronza da Silva	High	Unclear	High	Unclear	Unclear
Ganesh 2010	Low	Unclear	Unclear	Unclear	Unclear
Goncalves-Pereira 2017	Low	Low	Unclear	Low	Unclear
Goodman 2017	Unclear	High	Unclear	Unclear	Unclear
Guerchet 2013	Unclear	Unclear	Unclear	Unclear	Unclear
Hendrie 2018	Unclear	Unclear	Low	Low	Unclear
Huang 2014	Low	Low	High	Low	Unclear
Jia 2014	Low	Low	High	Low	Unclear
Khedr 2015	High	High	High	High	Unclear
Kim 2011	High	Low	Unclear	Low	Unclear
Koller 2010	Low	High	Unclear	Unclear	Unclear
Nadel 2014	High	High	Unclear	Unclear	Unclear
Nunes 2010	High	Low	High	Low	Unclear
Pilleron 2015	Low	Low	High	High	Unclear
Poddar 2011	High	High	High	Unclear	Unclear
Prince 2012	Low	High	High	Low	Low
Raina 2014	High	High	High	Low	Unclear
Rodriguez-Agudelo 2011	Low	Unclear	High	Unclear	Low
Samba 2015	Unclear	High	Unclear	Unclear	Unclear
Sharifi 2016	Low	Low	High	Low	Unclear
Taylor 2017	Low	High	High	High	Unclear
Tengku Aizan 2010	Unclear	Low	High	Unclear	Unclear
Tola-Arribas 2013	Low	Low	Low	Low	Low
Tornau 2015	High	Unclear	High	Unclear	Unclear
Tripathi 2012	High	Low	Unclear	Low	Unclear
Weden 2018	Low	Low	High	Unclear	Low
Wu 2017	Unclear	Low	High	Low	Low
Yin 2016	Unclear	Low	High	High	Low
Yuan 2016	High	High	Unclear	Low	Unclear
Zhao 2010	Low	Low	High	Low	Unclear

Study

D1: Selection of participants
D2: Confounding variables
D3: Measurement of exposure
D4: Incomplete outcome data
D5: Selective outcome reporting

Judgement
 High
 Unclear
 Low
 Not applicable

2.4 How does this study contribute to existing literature?

In the manuscript in Section 2.3 ([Published manuscript on *medRxiv*](#)) I outline how my research contributes to existing literature, highlighting a previous systematic review undertaken in 2012 [340]. My systematic review differed from the results of this previous one but with some important distinctions that might influence why the results suggested a different direction of association (increased dementia in urban areas in my systematic review compared with increased dementia in rural areas in the previous one) [1, 340]. Overall, geographical variation of dementia within countries exists but is not fully explained.

This previous systematic review [340] found 25 studies (by the search date of April 2010) to have investigated rural/urban comparisons in relation to dementia prevalence or incidence. Meta-analysis showed there was evidence of an association between rurality and prevalence of Alzheimer disease dementia (OR= 1.50, 90% CI 1.33-1.69), but less evidence for an association between rurality and non-specific dementia prevalence (OR= 0.91, 90% CI 0.645-1.83) [340]. Overall, for all prevalence studies, there was an intermediate risk of dementia for those in rural areas, regardless of sub-type of dementia, (OR= 1.11, 90% CI 0.79-1.57) [340]. For rurality and dementia incidence there was also an association with rurality (OR= 1.20, 90% CI 0.84-1.71) [340].

Additionally, a systematic review in 2000 investigated the relationship between education and place of residence and risk factors for Alzheimer disease dementia in African Americans in a USA city [235]. This systematic review found that childhood rural residence, combined with low level of education, was associated with increased risk of Alzheimer disease dementia. They found that odds ratios were 6.5 (95% CI:2.6-16.7) for low education/rural residence and risk of Alzheimer disease dementia compared with odds ratio of 0.5 (95% CI: 0.1-2.9) low education

and urban residence [235]. This contrasted with an odds ratio of 1.5 (95%CI: 0.4-5.2) high education and rural residence (high education and urban residence was the reference) [235]. The study suggested that low levels of education alone may not be a major risk factor but a reflection of other factors such as socioeconomic or environmental influences in childhood and future Alzheimer disease dementia risk [235].

It is known that there are differences in dementia prevalence between rural and urban areas. This is as yet, not fully explained. Some studies show higher rates in urban areas, or lower rates in rural areas (whilst recognising potentially undiagnosed people with dementia in rural areas) [343]. Other studies show higher rates of dementia in rural areas [344, 345]. One suggestion from these studies that the rural rates of dementia are twice that of urban locations could be attributable to larger populations of older people in rural areas [344, 345, 346, 347]. Another study suggested that those living in rural communities were at increased risk of dementia parallels other *"persistent place-based health disparities"* [345].

It is complex as one of these studies also highlighted a disparity in risk-adjusted Alzheimer dementia and related disorders (ADRD) whereby diagnostic incidence was higher in rural versus metropolitan areas despite lower prevalence in rural areas; concluding that older adults in rural communities may be under-diagnosed with ADRD or diagnosed at later stages of dementia [344]. Conversely, a study from Sweden in 2019 suggested that those living in rural areas with dementia were more likely to receive *"a complete basic examination, MMSE examination, Clock test, blood analysis, and neuro-imaging, compared to patients living in urban areas"* [348].

Interestingly, 2023 study investigating differences in the measurement of cognition for the assessment of dementia cross geographical contexts, highlighted that several studies had found that urbanicity could *"impact performance on cognitive items, holding underlying cognitive ability constant"* [349]. The study aimed to

"provide concrete guidance for dementia measurement in future Cross-national efforts" by using the Harmonized Cognitive Assessment Protocol (HCAP) surveys [349].

Overall, research, including work since my systematic review, has investigated geographical variation of dementia in general as well as within specific countries and comparing specific locations (such as rural/urban). Ultimately, whilst this highlights that there is variation between rural and urban areas, it also highlights that it is not yet known how this is translatable between countries and between different periods of time.

A number of studies have investigated dementia incidence or prevalence variation geographically between different geographical areas, whilst others solely look at specific dementia related aspects topics including dementia mortality, dementia care or anti-dementia drugs prescribing [350, 351, 340, 352, 353, 354, 355, 356, 357, 358]. Other studies consider solely the role of rurality and/or urbanicity in the risk of dementia, or to explain the variation of in dementia or aspects related to living with dementia [359, 360, 236, 361, 359, 362, 363, 348, 364, 345, 365, 366].

Research has investigated non-random geographical variation of dementia within mono-zygotic twins and highlighting the question of nature versus nurture in life long risk of dementia [350]. As outlined in Section 2.3 (Published manuscript on *medRxiv*), rural living in Scotland and Sweden were associated with an almost two-fold increased risk of dementia in Northern versus Southern parts of the countries.

Conversely, in the USA increased risk of dementia was associated with Southern versus Northern areas of the country. One study in 2023 found the prevalence of dementia varies from data between 2000-2016 from the USA Health and Retirement Study [367]. Out of $n=96,848$ observations, the highest rates of dementia

were in the South and the lowest were in the North-east and Midwest USA [367]. Southern birth was significantly associated with dementia and greatest risk for Black people and older adults with less education [367]. The authors also found that socio-demographic disparities were largest if residing or born in the Southern states of the USA [367]. Unfortunately, these studies do not investigate rurality or urbanicity in the location and risk of dementia.

Location of birth as a risk factor for dementia has mixed results in the USA. One study, showed that dementia mortality was higher if born in the Southern States of the USA (in the "*Stroke belt*") and this risk was irrespective of place of death [368]. Some studies, including two since my systematic review search date, have shown place of birth (Southern USA) was independently associated with a diagnosis of dementia and reduced cognitive functioning in older adults [369, 370, 371]. Whilst other studies in the USA have found that geographical variation of dementia mortality by birthplace can be explained by an individuals geographical residence in old age [372, 373]. Again, these studies have looked at variation geographically by North/South rather than rural or urban location.

From a more clinical perspective looking at dementia risk in studies using general practice data with a geographical (spatial) analysis, a systematic review in 2017 investigated this [352]. Of the $n=1,922$ citations screened in the search strategy, fourteen studies were included [352]. Three of these studies showed geographical associations and one showed no clear geographical variation in dementia risk distribution. Meta-analysis was not undertaken [352].

In summary, there is a growing body of work, and increasing research in this area, considering the role that the environment and geographical location has on the risk of dementia. Research has investigated variation of dementia at different geographical scales including North versus South variation and rural versus urban living. My systematic review, as outlined in the manuscript, builds on this by confirming overall that rural living (compared to urban living) is associated with

a small increased odds of dementia (OR= 1.20, 95%CI 1.03-1.40). It also demonstrates for the first time, as far as I am aware, that this is driven by increased odds of dementia in rural areas in Low to Middle-Income Countries (L-MICs) rather than High Income Countries (HICs). Importantly, it further highlights a difference between HIC and L-MICs showing that the odds of dementia were reversed in HIC with higher odds of dementia in urban areas in large studies (routinely collected data studies) in HICs.

2.5 Further discussion since the systematic review

In this section, I discuss key themes that have emerged from the systematic review, expanding the discussion that is already included in the manuscript. These themes are relevant as further discussion of the results of my systematic review and meta-analysis, and also for future direction of research in this area. Additionally, I discuss some papers published since my systematic review of relevance to expanding the discussion of this topic. In the subsequent Section 2.6 ([Future work](#)), I discuss future plans for updating the systematic review following changes in search strategies for geographical systematic reviews.

2.5.1 Defining an area or ‘place’

The definition of what we refer to as a ‘place’ geographically can change over time and also between regions and countries within the same time period. Previous publications have highlighted the complexity of investigating the role of a place or area and subsequent risk of dementia when definitions of a geographical area can be very heterogeneous, or individuals attachment to stay in an area varies in different cultures [[374](#), [268](#), [340](#)].

The definition of ‘place’ can mean many things when investigating location in relation to a disease trajectory or risk/protective factors for a condition. There are the physical boundaries of that area geographically, the level of rurality or urbanicity which in itself can change and be hard to quantify meaningfully in a

way that reflects variables or confounders that might influence interpretation of overall results.

One review highlights this need for a "*clearer definition of 'place'*" [268]. Indeed, it further discusses the need to encompass parameters within that 'place' which might be influencing cognitive decline or reserve such as information on neighbourhoods, communities and cities, broader spaces and other variables that might influence an older person's ability to carry out their activities of daily living [268]. As there is growing interest in 'cognitive-friendly environments' and environments which nurture 'brain-training' or cognitive reserve [374], it is important that 'place' is not considered as a binary variable in terms of rural/urban but reflects the complex structure in which an individual lives as well as how that might change across their lifetime.

As highlighted in [Chapter 1: Introduction](#) Section 1.5.2 ([Location: geography and brain health](#)) the definition of rural and urban has also undergone much discussion at an international level, including within the United Nations (UN), with a move to a unified definition on what is classed as urban versus rural, which is also translatable across countries [375, 376, 377]. Therefore, there have been key developments in this area since my systematic review, although not yet translated to being used consistently within research in this field.

When considering the geographical variation of a disease such as dementia, it is important to recognise the longitudinal aspect and how changes in areas may be defined both in terms of boundaries but also in terms of facilities present within that environment. All these aspects (or features of what defines an area) can change with time. This is an ongoing challenge for research investigating the geographical variation of dementia at many levels: defining a rural and urban setting, considering the features of an environment that influence an individual's health and well-being, how this changes over time, and lastly the underlying influence that the environment may have on our genetics influencing our risk of dementia

is discussed via epigenetics in Section 1 ([Location, ageing and cognition](#)). It is a complex, growing and fluid area of research to investigate the role of where we live geographically and our risk of dementia.

2.5.2 Location (rural/urban) across our life-course as a risk factor in dementia

The risk of Late onset dementia increases with age, with possible accumulation of risk factors or certain ‘sensitive’ or ‘critical’ time points in a person’s life influencing a future diagnosis of dementia. Where we live has an impact on our health, well-being and risks of disease with a complex interplay between an individual and the environment in the location they live. What we term as ‘place’ and what comprises that geographical area consists of more than solely geography and physical environment but the social environment that is intricately connected. Since, my systematic review search (2018), there has been further research investigating rural or urban location and dementia outcome which I will outline in this section.

Access to services is an important aspect to consider within an environment that might vary between different rural or urban areas within the one country and across a person’s life-course impacting on potentially modifiable risk factors being (or not being) treated appropriately. This again reinforces the need to understand ‘place’ when defined in a geographical context for dementia epidemiology research. Different countries have their different methods for categorising geographical areas, often due to population size and density. This will be discussed further in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#) in relation to Scotland’s geographical scales and definitions.

Of importance, even if rural and urban definitions are defined in the publications, there is an evolving situation in both HICs and L-MICs in the population dynamics and infrastructure of urban areas across an individual’s life-course. Rurality

and urbanicity are dynamic constructs. In Scotland, the movement of individuals to rural areas is increasing. However, globally urbanisation is expanding both in HICs and L-MICs, but greatest to be projected in L-MICs. In L-MICs the projected urban expansion between 2000-2050 reflects urban population growth and density change [378]. The world's urban land cover will increase to 114% between 2000 and 2050 as the world population doubles [378].

The paper in 2011 (prior to the global definition of rural and urban) highlighted that there was no reliable global map to identify all land in urban use and that without that, inability to estimate the amount of urban land in different countries or impact from the expanding worldwide population size [378]. When location is dynamic across a person's life-course then the potentially modifiable risk factors (or protective risk factors) within that location impacting on the individual are varying. The temporal and spatial elements of potentially modifiable risk factors within a location need considered across an individual's life-course for dementia risk.

A 2019 systematic review specifically investigated early-life risk factors for dementia and cognitive impairments in later life [211]. This found that, in a prospective study ($n=2,711$), childhood rural residence (before the age of 19 years old) could increase the risk of dementia by 52% compared with urban residence [211]. This was highlighted in Chapter 1 [Location: geography and brain health](#). Overall there are few studies that investigate location in terms of rural and urban location in childhood as well as across an individual's life course with their future dementia risk. This is not unsurprising given the complexity of investigating location, and the many parameters that contribute to location, at one point in time as well as at multiple points across an individual's life time in the context of dementia risk.

A life-course approach is challenging in designing and initiating a study which is longitudinal and captures this information. Cohort studies are a resource to do this, and are discussed in [Chapter 3: Using big data for dementia research](#),

when *nested* or linked to routinely collected health data (with geographical data). The potential of unique geographical identifiers integrated into routinely collected health data in the future, such as the Unique Property Reference Number (UPRN) [379], in the UK may provide a valuable opportunity to do this within the Five-Safes Framework (Table 3.2) and across decades for a longitudinal approach which could allow a systematic review focused on UK studies to understand rural/urban variation of dementia across the four nations.

There is additional complexity of investigating rural and urban living and dementia risk across different countries at different points in time (or an individual's life-course) with recent differences in how individual countries define rural and urban areas. Since my systematic review, there have been other country specific systematic reviews looking at rural and urban variation of dementia. Whilst the urban and rural areas can change across time within a country, removing comparison with other countries reduces the heterogeneity in the methods between studies. One recent (2024) systematic review and meta-analysis investigated the epidemiology of dementia in China over a 10 year period (2010-2020) [380].

This systematic review found that in sixteen of the nineteen studies included in the meta-analysis, the prevalence of dementia was 6% (95%CI 5%, 8%; $I^2 = 99\%$, $p < 0.001$) [380]. Of relevance for this discussion, there was no statistically significant difference between the prevalence of dementia in rural and urban areas [380]. Indeed, the sub-group analysis of different areas found the prevalence of dementia in rural areas was 6% (95%CI 4%, 8%; $I^2 = 87\%$, $p < 0.001$), and comparable to the prevalence in urban areas 6% (95%CI 4%, 8%; $I^2 = 98\%$, $p < 0.001$) [380]. Whilst there was no clear difference, interestingly they found the prevalence of dementia decreased by year in urban areas but not in rural areas [380].

Conversely, another systematic review in 2022 in Latin America Caribbean (LAC) countries found a difference between rural and urban areas for dementia preva-

lence. The study found prevalence of all-cause dementia was 8.68% for rural areas (95% CI: 4.74–13.65%) and 7.71% for urban areas (95% CI: 6.43–9.09%) amongst $n=96,396$ participants in their meta-analysis [381]. We do not know how research on location and dementia risk in HICs translates to L-MICs as the potential risk factors in different areas may change with time and space.

Many of the risk factors (or protective factors) for dementia that could vary between location across an individual’s life-course have been predominantly studied in HICs. Changes in the structure and function of rural and urban locations have changed across the 80 year life-course in a HIC of an individual’s life for those now at greatest risk of dementia due to their age. However, this is not necessarily translatable to different cultures, contexts and countries.

Whilst urban environments can provide many benefits such as potentially providing better access to health care, employment and education [382] which may subsequently impact on an individual’s life-course risk of dementia, living in cities may also be detrimental to mental health outcomes. This could be attributable to individual, social and environmental factors across an individual’s life-course and subsequently influence their risk of dementia depending on location [383].

One such variation between rural and urban areas that can influence physical and mental health and well-being is access to green-spaces. A recent scoping review [300] explored the uncharted territory of whether urban green-spaces support mental health in low-and middle-income countries. The authors undertook a scoping review to assess the extent, type and quality of evidence contained in the heterogeneous body of literature investigating the impact of green-spaces on mental health in L-MICs [300]. The review found significant heterogeneity and was unable to undertake meta-analysis [300].

This narrative review found that of the included 36 studies, 31 highlighted one or more positive health promoting relationships between green-spaces and a mental

health outcome(s) [300]. The review highlighted that the relationship between green-spaces and mental health remained unexplored in more than 90% of L-MICs [300]. This is just one of many possible differences between individuals living in rural and urban locations and the impact on dementia risk, but highlights that there is still much research needed globally to understand the variations of this and then subsequently understand the life-course implications of variables changing within a location to influence rural or urban differences in dementia risk across an individual's life-course.

In summary, understanding how the environment in which a person lives, and what association there is between the environment and the individual, is important within this model of causality for a life-course dementia risk. It is also important to contextualise this as comparisons between countries can be helpful and sometimes unhelpful. Rurality and urbanicity are dynamic constructs that change with time, within and between countries. This creates additional complexity across an individual's life-course epidemiologically.

2.5.3 Location as a risk factor for dementia in different country contexts

HICs versus L-MICs

It is already known that there are reported differences in dementia incidence and prevalence between countries [92]. A recent systematic review (2024) [384] investigated this from analysing cohort studies. Europe ($n=5$) and USA ($n=5$) studies consistently reported declining incidence in dementia. The Population Attributable Fractions (PAFs) for less education, smoking, or both declined with time, and within the Framingham Study (USA 1997-2003) [91]. These were associated with a decline in incidence of dementia [91].

As discussed in [Chapter 1: Introduction](#), future worldwide incidence and prevalence of dementia is anticipated to have the greatest increase in Low and Middle

Income Countries (L-MICs). The World Health Organization (WHO)’s report, ‘Global status report on the public health response to dementia’, states that in 2021 of the 55.2 million people worldwide living with dementia 60% live in LMICs [385]. In this sub-section I discuss three aspects of relevance to consider the different country contexts when exploring location (rural and urban living) as a risk factor for dementia.

In my systematic review there was an important finding, which I believe has not been previously published: *rural living was linked with an increased odds ratio of dementia in L-MICs but this finding was reversed in HICs*. I also highlighted that this difference was most evident when considering large data using routinely collected data in HICs versus L-MICs.

As highlighted in the manuscript, I believe this has importance for future research in this field as the use of routinely collected data for research in HICs is increasing, but not at the same pace in L-MICs [386, 387, 388]. One limitation of the systematic review was the high heterogeneity in the studies, including the methodologies. Studies using routinely collected data were predominately done in HICs. The size of these studies are naturally much larger as they are using big data (thousands to hundreds of thousands of individuals) versus traditional face-to-face epidemiological studies that have much smaller sample sizes.

Data challenges (using routinely collected patient/health data or disease registries) have been highlighted in other diseases in L-MICS [389, 390, 391, 392, 393]. Whilst some L-MICs have national registries, disease specific databases, routinely collected health data or administrative routinely collected data, there are common similarities reported regarding the data challenges. These include that these resources are underused for understanding disease prevalence in L-MICs, and critically that there are logistical challenges of development and widespread implantation of using routinely collected data in L-MICs [393, 391].

Traditionally, some of the data collection was focused specifically on communicable diseases. However, with the increasing recognition of the importance of non-communicable diseases (NCDs) (the prevention and management) this is changing in some countries [391, 393]. A recent (2022) scoping review in China established 70 databases that were useful for health data research [391]. Of these, only 20 were for non-communicable diseases [391]. The authors highlighted that with <30% of routinely collected databases related to non-communicable diseases there are challenges that researchers are trying to navigate [391].

China has more databases compared with some L-MICs, but not compared with other countries in Asia-Pacific. A review in 2012 investigated data resources for health research in Asia-Pacific countries (Thailand, China Mainland, South Korea, Taiwan, Japan, and Malaysia) and found that there was a wide range of healthcare data sets which were routinely collected (and additionally other sources such as national surveys) but that this depended on the economy's perspective [394]. One important concern for use of routinely collected health data for future research in L-MICs investigating dementia was noted in this review: *"Another concern is the accuracy in the coding of diseases and procedures, which is often done by health staff with little or no training in coding"* [394].

In HICs, we know that dementia ascertainment can be obtained from routinely collected data of ICD codes or anti-dementia drugs. We don't know how this translates to routinely collected data in L-MICs. This would result in similar methodological papers and less heterogeneity in a future systematic review (if more L-MICS used routinely collected data for geographical variation of dementia), but could introduce additional bias in interpreting the results. Unlike traditional face to face epidemiological studies which clearly define how dementia is diagnosed, with routinely collected health data we do not know the process that came before the ICD code was documented. Additionally, in HICs we know that the sub-type of dementia is not always captured well in routinely collected health data.

The Global Dementia Prevention Program (GloDePP) collaboration met in Edinburgh, and in their subsequent review in 2019 highlighted a strategic goal: *"Develop a "global data science platform" to enable data sharing from diverse sources that could be used to study dementia in greater detail especially focusing on cognitive decline across different groups and contexts globally"* [395]. A further review in 2019 highlighted the potential of L-MICs to use big data and Artificial Intelligence (AI) in the future for dementia research in a collaborative approach with HICs to increase understanding, and prevention, of dementia. This is important but may be an ambitious immediate goal given AI for dementia research in big data is in the early stages in HICs where routinely collected health data is more regularly used [396].

The comparison, as highlighted in the systematic review manuscript, between HICs and L-MICs has additional complexity given that a number of countries have changed their classification by the World Bank in recent years. For instance, Malaysia (now classed as an Upper Middle Income Country by the World Bank) has an extensive routinely collected health data resource. Research from 2020 in Malaysia using a nationwide survey ($n = 3,774$) found dementia prevalence was highest in rural areas (12.9%, 95%CI 10.50-15.84) versus urban areas 6.8% (95% CI 5.11-9.00) [397]. This study also found that prevalence decreased with increasing education level and it was unclear the differences in these between the rural and urban areas [397].

This emphasises the point I highlighted in Section 2.5.1 Defining an area or 'place' that defining an area is important, but that this can change over time. This could be more pronounced in countries where they are climbing the economic ladder now, but the population at risk of dementia have experienced a different economic landscape in their childhood and early adulthood. This adds complexity when comparing rural and urban locations and the socio-economic environments within those locations (including access to healthcare services or education) which

could create bias and needs careful consideration when interpreting variation in rural/urban locations between HICs and L-MICs. We know that, to date, dementia risk models have exclusively been developed and tested in HICs [398], despite most people with dementia living in L-MICs, and only one has considered rural location in its model.

In summary, there are important differences between countries (particularly HICs and L-MICs). These include growing disparities in the use/non use of routinely collected data between the two groups for investigating dementia and rural/urban differences. Additionally, the factors that contribute to rural and urban environments in different economically impacted countries could create different potential risk or modifiable factors that are driving differences that I found in my systematic review.

Lastly, the global environment is changing with increasing urbanisation (and the potential increased future urbanisation in L-MICs). When analysing the results of my systematic review it is important to consider temporal change in the spatial variable (rural and urban environments). Ideally, capturing rural/urban location over several time points of a person's life with clear definitions of what constitutes rural or urban at those points of time is needed. However, this is challenging and none of my studies included these factors. Standardisation of rural and urban definitions will be one step to begin to improve this.

Multiple ethnicities and indigenous populations

Another important aspect that arises from the need to consider the variations between countries and conclusions drawn from my systematic reviews is the need to recognise that many studies do not include multiple ethnicities or indigenous populations. We do not know how a study undertaken in one country with one ethnicity translates to another country with a different ethnicity. Drawing conclusions from my results therefore needs a cautious approach.

There may be disparate health inequalities that might be influencing variation in dementia risk by location (rural/urban living). My systematic review did not capture ethnicity and this is something I am amending in my updated systematic review, particularly in view of my results (as well as learning and experience) in Aotearoa New Zealand in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#). In Scotland, my research in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) is on a population in 1947 on people with white ethnicity. We know that the majority of dementia research occurs in HICs, of that most is focusing on white ethnicities although there has been important work in USA investigating more ethnicities.

It is important to be inclusive of all ethnicities and when cohorts or datasets are created that (due to many factors) have predominately one ethnicity, it is important to recognise that the results are not necessarily relatable to different countries and ethnicities [\[399\]](#). This may be even more important for my thesis topic given the potential role that the gene-environment may play on epigenetics and also health disparities within different areas that may impact differently on multiple ethnicities.

A study from Australia investigating rural and urban dementia diagnosis in indigenous and non-indigenous populations found $n=784$ diagnosed with dementia and $n=249$ estimated undiagnosed with dementia [\[400\]](#). In summary, the study concluded: *“Undetected cases in the Indigenous population were over-represented for the population, with 47.8% of the estimate of unknown cases being Indigenous, whereas 30% of total NT [Northern Territory] population are Indigenous”* [\[400\]](#). which the authors concluded was consistent with lower access or use of services in remote areas [\[401\]](#).

In the UK, as in many countries, we know that there are differences and barriers to accessing dementia services for ethnic minorities [\[402, 403\]](#). This is despite higher risk of dementia in Black ethnic groups [\[404, 405\]](#) and increased co-production

of dementia research that is culturally and ethnically diverse is focusing on this [406, 407].

Considering this in the context of my systematic review, the studies included in were predominately investigating dementia prevalence. A systematic review in 2017 on dementia prevalence and incidence in United States in different races and ethnic populations highlighted a focus that prevalence comparisons can introduce bias [408]. This systematic review highlighted that incidence rates were preferable to avoid differences in lengths of illness amongst racial and ethnic populations due to different ages of onset and life expectancies which comparisons of prevalence studies can miss [408]. My updated systematic review will draw reference to these important aspects when interpreting my findings.

In summary, throughout my thesis I highlight the need to generate country-specific population level (with individual level detail) dementia data to be able to then address the question of how location (rural or urban living) influences dementia risk. This is important in both HICs and L-MICs. As a 2024 publication from the 2022 symposium on dementia and brain ageing in low- and middle-income countries highlighted: *"Two-thirds of persons with dementia live in L-MICs, yet research and costs are skewed toward HICs"* [409].

It is difficult to understand the reasons explaining the direction of risk in my systematic review (increased odds of dementia in rural areas in L-MICs compared with increased odds of dementia in urban areas in HICs). It does highlight this is important to further explore. In particular, will this result be similar in my future updated systematic review, and if so what might be driving this. There is a need to explore how rural and urban areas differ between HICs and L-MICs, and across an individual's life-course to understand the risk of 'location' for dementia.

2.6 Future work

In this section I discuss some of the changes in systematic review searches that are being developed for improving geographic filters that are relevant as I update my manuscript for submission to a peer-reviewed journal. Additionally, I comment on some of the feedback from the editor of a peer-reviewed journal we submitted to and include some additional discussion in this section relevant to that which will expand my revised systematic review.

2.6.1 Geographic search filters

My systematic review search date was 31st October 2018. Recently there have been major developments in trying to understand the optimum geographical filters for systematic review searches on geographical topics [410]. This is an important development.

Geographic search filters are applied with the aim to retrieve evidence or publications related to those geographical locations (such as continents or countries). As of 2020, there were only three validated geographical filters available in published literature covering Spain, Africa and the UK [411]: 1) Spain: PubMed [412]; 2) Africa: PubMed and Embase [413]; 3) UK: MEDLINE and Embase, OVID platform [414, 415, 416, 417, 418, 419]. For L-MICs there are limited geographic filters for search terms as highlighted in a recent (2022) paper [420].

My systematic review search was 2018, prior to the evolving discussion around reliable geographic filters. I worked with the information scientist, and to ensure we did not exclude any countries (or variations of country names) we included multiple options for countries. This is highlighted in Subsection 2.3.1. This was laborious and takes time. This highlights how important it is to have reliable and consistent geographic filters for systematic review searches to avoid inter-research variation in searches.

The UK National Institute of Clinical Excellence (NICE) guidelines search within OECD (Organisation for economic co-operation and development) countries in developing their recommendations [418, 419]. It is important to have clearly defined geographic filters so information is not missed, nor laborious time wasted listing every version of a country.

Previous work in Germany highlighted the variation in terms used when searching for articles from Germany [421]. This review identified that search strategies revealed a need for improvement for a regional search filter for Germany to identify studies performed in Germany. There was variation in the use of the terms used to describe a study in Germany such as Germany, German or Deutschland [421]. There is relevance of this to other regions to ensure all relevant publications are captured for a systematic review search with a geographical focus, whilst not creating extensive search lists without using multiple different terms possible for each country [421].

The NICE UK filters for MEDLINE and Embase were previously developed in 2017 and 2019 [414, 418, 419, 415, 417]. In 2021, two studies first developed the methodology and then assessed the validation of the NICE OECD countries' geographic filters in MEDLINE and Embase (Ovid) search engines [418]. This study found that the geographical filters both achieved a 99.95% recall against the gold standard sets and when validating the filters against three NICE guidelines reviews achieved a 100% recall [418]. Additionally, it reduced the number needed to read (NNR) on the search from $n = 256$ to $n = 232$ for one review, from $n = 38$ to $n = 27$ for a second review, and from $n = 631$ to $n = 591$ for a third review in MEDLINE. There was a similar reduction in NNR from using these geographic search filters in the Embase search engine [418].

A recent publication in 2024 of a scoping review protocol for investigating geographic filters aims to *"identify existing validated geographic search filters and*

summarize the development and validation methods for each of them. In addition, we will report the retrieval performance of the filters in terms of sensitivity (also known as recall), precision, specificity, and number needed to read (NNR)" [416].

In view of this development progressing I am updating my systematic review with the new recommended geographic filters. Our original search was conducted by an information scientist following numerous meetings between us. To avoid excluding papers, our search was broad and used all country names to try to minimise us restricting our search too much and inadvertently excluding any relevant papers. This is very much a developing and expanding field in the last few years and likely to become more precise following the results of the scoping review mentioned above.

Working with an Information Scientist (IS) aware of these developments is really important and this is underway with the IS who was co-author on the above manuscript. This discussion highlights that for researchers to be able to produce evidence based research on health inequalities between regions and geographical areas there needs to be assurance that searches will rightly recognise all publications that need considered in a systematic review. It feeds into a wider picture, about how a location is defined which I discussed in the manuscript, and Section 2.5.1 ([Defining an area or ‘place’](#)).

2.6.2 Heterogeneity of studies

In my systematic review there was substantial heterogeneity amongst the studies included. This was discussed in the manuscript. For the purpose of this systematic review, an I^2 value was considered heterogeneous when over 50% [422]. The I^2 describes the percentage of the variability in effect estimates due to heterogeneity rather than sampling error (chance).

Much of the heterogeneity in the studies in my systematic review were around: 1) the outcome (dementia) ascertainment and differing methods used, 2) variation

in how a rural or urban area was defined, or not defined at all, 3) the study design such as a traditional face-to-face epidemiological observational study versus using routinely collected data to undertake epidemiological observational research. As I update my systematic review with a revised search date and updated geographic filters, I will be collecting further information on these three aspects of the current systematic review that showed much heterogeneity with the aim to be able to undertake sub-group analysis in the meta-analysis with studies that are more homogeneous in these different areas, particularly between the study design.

It is not unsurprising that there was a lot of heterogeneity in the studies in my systematic review. Interestingly, a recent systematic review in 2022 [381] investigating the prevalence of dementia in Latin America and Caribbean countries focusing on rural and urban variation also found high heterogeneity between studies.

Heterogeneity can be in different contexts: clinical heterogeneity, methodological heterogeneity, statistical heterogeneity. For the updated systematic review, these types of heterogeneity will be more clearly summarised with key reasons studies fall in the different types of heterogeneity outlined. Additionally, I will further explore the heterogeneity in each of these groups by conducting further subgroup analysis. For instance, there was a lot of clinical heterogeneity in the methods used for ascertaining or diagnosis dementia. Different methods will be grouped together and analysed as sensitivity analysis in the meta-analysis.

There was methodological heterogeneity in the type of study design (routinely collected data versus traditional face-to-face epidemiological study designs). These will be sub-grouped and analysed separately. There was also methodological heterogeneity in the definition (or lack of) for rural/urban location and also whether it was in a HIC or L-MIC. I did sensitivity analysis on the latter. For statistical heterogeneity, I used random-effects model to incorporate heterogeneity among the studies

For my updated systematic review, a treemap of the most commonly used variables (predictors of dementia), with size and colour representing frequency and variable category will also be included alongside the exposure of interest (location: rural or urban).

In summary I recognised there was heterogeneity of studies in my systematic review. I commented on this, and adjusted to assess for this, but the updated systematic review will include additional sub-group analysis to account for the heterogeneity, and assess if this influences the overall findings.

2.6.3 Risk of bias assessment

Risk of bias (ROB) is defined as: "*a systematic error, or deviation from the truth, in results*" by The Cochrane Handbook for Systematic Reviews. It clearly highlights that this should not be confused with *imprecision* (random error) or *external validity* (generalisability) of the study [423]. Risk of bias can also include non-inclusion of studies, for instance if incorrect terms are used in searches. This is important given my search strategy and subsequent changes in geographic filters for searching in medical search engines which was highlighted in Subsection 2.6.1.

Undertaking a systematic review can result in risk of bias in multiple stages: risk of bias in results of included studies, risk of bias due to missing results and risk of bias due to conflicts of interest of study investigators or funders [423]. For systematic reviews on epidemiological studies (non randomized studies) selection bias and information bias are more of a concern than reporting (publication) bias.

As part of my systematic review I use the terms risk of bias assessment and the ‘quality assessment’ in the same way. I appreciate that risk of bias is a primary driver of the quality of a study, but quality assessment can encompass a wider quality assessment. I adapted the The Risk Of Bias In Non-randomized Studies of Interventions (ROBINS) tool for my systematic review [424, 425, 426].

This tool focuses on a domain-based (or component) approach, where different types of bias are considered individually [424]. My tool followed the same format as recommended, assessing seven domains: *"random sequence generation, allocation sequence concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other sources of bias"*, with no scoring system [424].

Additionally, there were signalling questions, as recommended, aiming to *"elicit information relevant to the risk-of-bias judgement for the domain"* [424]. As per recommendation, my ROB assessment tool required authors to judge each domain as 'low', 'high' or 'unclear' risk of bias. I did not capture all the reasons provided to support their judgement in the manuscript but will be adding this information in the updated systematic review.

Two people reviewed each paper for assessing risk of bias in my systematic review, following the Cochrane recommendations [424]. The risk of bias assessment tool (quality assessment) form was created through discussion with my PhD supervisors, and co-authors, (Dr Tom Russ and Professor Susan Shenkin) to ensure expertise in the area. The form used is in Section 2.3.1 ([Systematic review appendices](#)). I also trialled this risk of bias assessment tool with both me and my principal supervisor to improve reliability of assessments. We had a clearly defined process for resolving any disagreements which was decided in advance of starting the systematic review.

There are other different quality of assessment options for assessing risk of bias in observational studies, such as the NOS (Newcastle-Ottawa epidemiological observation studies risk of bias) [427]. All have positives and negatives as outlined in a summary review paper [425, 428]. Importantly, this year (2024) a new ROB tool, the Risk of Bias in Non-randomized Studies - of Exposure (ROBINS-E) was created [429]. ROBINS-E tool provides a structured approach to assessing the risk of bias in observational epidemiological studies investigating exposure

(e.g. environmental). My updated systematic review will aim to use this new tool (ROBINS-E) in addition to the ROBINS [425] tool used for the systematic review in this chapter. This will allow consideration of different assessments of ROB, and also a consistent comparison between my results in this systematic review and updated systematic review. It is recognised that there is poor inter-rater reliability of all ROB assessments.

The striking result in my risk of bias assessment was that the definition of rural/urban location (exposure) was exceptionally high. The international variation in the how a rural or urban area is defined means these results whilst striking were not surprising. It also highlights complexities in drawing conclusions from different countries (which define rural/urban differently) and how this translates to other settings. A number of studies did not define rural or urban in their research in any form.

There are three recommendations from Cochrane in considering risk of bias in the analysis [424]. Of these, the first suggests restricting the analysis to studies at low risk of bias [424]. The second recommendation is to present multiple (stratified) analyses [424]. The third recommendation is to present all the studies and provide a narrative discussion of the risk of bias [424]. I selected the third recommendation as this had been used previously in the older systematic review, and also because exposure (geographical - rural/urban location) had high risk of bias across most studies. As there was often no clear definition of rural or urban locations in the studies leading to bias it was important to provide a narrative discussion of the risk of bias in relation to this.

I addressed this in the discussion in relation to the high risk of bias in the exposure (geographical location) domain. For my updated systematic review I will expand this, and also collect further information on the detail of how the exposure varied. Additionally, I will include further sensitivity analysis to show whether the results changed when restricted to studies with low or no risk of bias in the classification

of exposure (in line with Cochrane’s first recommendation to undertake: *primary analysis restricted to studies at low risk of bias* [424]. I will include reasons for all the five domains of each study and not only the grading of that risk of bias.

2.6.4 Publication bias

Publication bias, or reporting bias, is due to studies with negative, or perceived to be uninteresting results, not being published. This can result in wrong interpretations in meta-analysis particularly if results with positive results, or results that are in keeping with expectations are more likely to be published and subsequently more likely to be included in meta-analyses. This can be an important bias in publication, or failure to publish, particularly in randomised controlled trials. Publication bias is classed as a type of *reporting bias*.

My systematic review did not discuss publication bias and so I discuss it in this section. In the The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach approach, observational studies are considered low-quality evidence compared with randomized trials which start as high-quality evidence [430, 431]. Both can be down-rated if a there is evidence of high risk of publications bias [430]. For observational studies, there are challenges in assessing publication bias.

A past review [432] discusses the approach to publication bias using two approaches: selection models, or funnel-plot-based methods. Whilst selection models use *weight functions to adjust the overall effect size estimate and are usually employed as sensitivity analyses* to assess the potential impact of publication bias, funnel-plot-based methods use a funnel plot (scatterplot) to visualise the presence or absence of publication bias. A funnel plot includes plotting the intervention effect sizes from the different studies (on the x-axis) against a measure of the study size (on the y-axis).

As the precision of the estimate of effect size increases with the study sample

size, smaller studies will be more widely scattered towards the bottom of the scatterplot. the variability will reduce as the study sizes increase. Publication bias, if present, would result in an asymmetric appearance of the funnel plot (a *unilateral gap towards the bottom of the funnel where the results of the small, negative, unpublished studies should have been* [424]).

An alternative approach for assessing publication bias would be to use Egger's test after the 'leave-one-out' method that another systematic review (with meta-analysis) used recently when assessing dementia prevalence in rural and urban areas in Latin America and Caribbean countries to assess for publication bias [381].

All these methods have difficulties in assessing publication bias in observational studies. For my updated search in my systematic review I aim to use a funnel plot to investigate publication bias.

2.6.5 Implications for further research

A key implication from my systematic review relevant for future research is the need to clearly define a rural and urban area, and also to have unity with this definition across different countries and regions. It is positive that there is now international uniformity in the definition of what constitutes an urban and rural area and that this can be translated across different countries. However, as with other universal agreements (such as the ICD-11 Version [21]) it can take time before these changes filter into being used consistently in research.

With the increasing use of routinely collected data for observational studies, another implication for future work is consideration of a recommendation for reporting clearly defined geographical data and location as an amendment to the REporting of studies Conducted using Observational Routinely-collected data (RECORD) guidelines [433]. This would allow a more uniform approach in

analysing studies using routinely collected data for this (and other) research questions in meta-analyses in the future.

Another implication of my systematic review for future research is that whilst those in L-MICs have the greatest potential increase of dementia in the future, L-MICs are more likely to have smaller traditional epidemiological face-to-face studies. There is work to increase the access and use of routinely collected data in L-MICs, but it is not universal [387, 386, 434, 388, 435, 436, 437]. Understanding the *selection bias* when creating a study population in routinely collected data is important, and how this varies from traditional epidemiological observational face-to-face studies is important when drawing conclusions from the results of meta-analysis. It is important to have an inclusive approach to dementia research, and not marginalise those in L-MICs as these countries have the greatest predicted increase of dementia prevalence over the next two decades.

A further implication is the recognition that most dementia studies are undertaken in HICs, and in general involve white populations. It is important to include ethnically diverse populations as the results from one country may not be translatable to other countries, ethnicities and cultures.

Lastly, an implication from the discussion in this chapter is the dynamic nature of what constitutes rural and urban locations across countries and across time. This is important when drawing conclusions on the results of meta-analysis of studies that involve different countries, different time periods and even if clear definitions of location are included as what constitutes that location may have changed since the publication date. Similarly, what constitutes a HIC or L-MICs can change over time.

2.7 Chapter conclusions

This chapter lays the foundation for my thesis investigating the role of rural/urban location and dementia risk. Since this systematic review, there has been further work discussing the role of the environment in relation to cognition and dementia, and research investigating geographical variation of dementia. The interest in green space, built environment and other measurable variables within a certain geographical area, alongside the impact on brain training, cognitive reserve is a growing field of research. Therefore, it is really important for future research, and understanding of factors that may be contributing to variation in dementia geographically, that clear definitions of a rural and urban area, that are translatable between countries, are now used.

Understanding the geographical impact on a disease such as dementia, which itself encompasses different sub-types with different aetiologies, is clearly a growing area of dementia research. As research increases in this area, the recognition of how an individual is integrally linked with their environment increases, particularly across their lifespan when the focus is on late onset dementia. In the future, ideally capturing within the rural/urban area what contributes to that in terms of environmental/geographical variables such as ‘green spaces’, ‘built environment’, ‘access to services’, ‘deprivation’ would be important. Currently, this is challenging with the breadth of variety of studies investigating environment or geography and a dementia outcome. These parameters might capture the ‘type’ of environment within the rural or urban location (and not solely population density or land mass). This may influence the overall results of the study in terms of quality and risk of confounders.

My systematic review also draws attention to the need to reflect on the growing use of large routinely datasets in health and administrative data to undertake research, and how this could create more heterogeneity in methodological techniques between High-, Middle- and Low-Income Countries.

In conclusion, my systematic review mirrored the previous systematic review highlighting similar heterogeneity of the studies, particularly related to the definition of rural or urban locations. My systematic review will hopefully highlight the need for clearer definitions and consensus agreements of geographical areas and locations that can be more constant over time and between countries. The developments of geographical filters for systematic review searches and ongoing developments (as discussed in Chapter 1) on the use of unique identifiers for locations in the UK that do not change over time may assist in reducing heterogeneity in methodologies and particularly the definition of rural and urban location which, as an exposure outcome, had high risk in the majority of the studies included in the systematic review.

Future studies need to carefully consider study setting, method of dementia ascertainment, when exposures may occur across a person's lifetime, and risk of bias, to understand the role of environment and geography in dementia risk and be able to draw meaningful conclusions.

Chapter 3:

Using big data for dementia research

3.1 Chapter introduction

In [Chapter 1: Introduction](#), I discussed the definitions of big data in healthcare and routinely collected data. As routinely collected data in healthcare are generated as a product of interactions of patients and healthcare services its primary purpose is to care for patients (allowing patients to be investigated, diagnosed and managed). This digital resource can be repurposed and have a valuable secondary use, which is for healthcare research [[155](#), [438](#), [152](#)]. Indeed, there has been a shift towards using ‘big data’ in the form of routinely collected data in healthcare research. In the prior [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#) I discussed the changing methodologies of investigating geographical variation of dementia, moving away from the traditional face-to-face epidemiological methods to interrogating large routinely collected datasets for research.

In this chapter I will build on the overview of big data and routinely collected health data in [Section 1.3 \(Overview of ‘big data’\)](#) with a specific focus on its use in research, and particularly highlighting how it can be used in dementia research. Alongside how it can be used for research, I discuss the challenges and processes to link and access data; the variability and reliability for using the data for dementia research; the UK landscape of healthcare data research; and the variation in accessing the large datasets both across the UK and internationally.

Finally, I will discuss the changing landscape of health data and health data research in the UK during my PhD, and how this might be further expanded specifically for research investigating key issues for older people such as demen-

tia. This chapter sets the context in which the analyses presented in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#), using big data research, are situated in order to answer my thesis question about how the role of location (rural and urban living), in relation to dementia diagnoses, can be investigated using routinely collected health data.

3.2 Peer-reviewed published manuscript

This chapter begins with my peer-reviewed publication [\[2\]](#) giving an overview of the big data landscape in Scotland as it was in my earlier stages of my PhD in 2017.

Walesby KE, Harrison JK, Russ TC. What big data could achieve in Scotland. *J R Coll Physicians Edinb* 2017; 47: 114-9. Copyright © [2017] (SAGE Journals). DOI: 10.4997/JRCPE.2017.201 [\[2\]](#).

Following the article I discuss in more detail how to re-purpose routinely collected data for healthcare research.

What big data could achieve in Scotland

KE Walesby¹, JK Harrison², TC Russ³**Declaration of interests:** No conflict of interests declared**Correspondence to:**KE Walesby
Alzheimer Scotland
Dementia Research Centre
7 George Square
University of Edinburgh
Edinburgh EH8 9JZ
UK**Email:**

katherine.walesby@ed.ac.uk

What are big data?

The term 'big data' is now frequently used in medicine. But what does this really mean and how is it relevant to the clinician as well as researcher? The background of this increasingly popular term is discussed here; what it means in clinical practice, for patients and for research; and why Scotland is poised to take advantage of this big data revolution.

There is no universal agreement on the meaning of big data. Indeed to many the term is vague, ill-defined and confusing. It generally refers to large or complex data, often requiring terabytes or petabytes of storage. It might refer to large amounts of information at a population, regional or local level or span different geographical areas. To put this in context, a terabyte could hold 1,000 copies of the *Encyclopaedia Britannica* and a petabyte 20 million four-door filing cabinets of information.¹

Traditionally, the term 'big data' has been linked to the explosion in the ability to capture and store data generated by the use of electronic devices in daily life. However, it is increasingly recognised that it also describes the large volume of patient information collected in everyday health and social care practice. Less well recognised is that using so-called big data for clinical improvement has a much longer history. In 1839, William Farr (a British epidemiologist) realised the benefits of routinely collected data when he established a system for recording the causes of death. This allowed mortality rates to be compared for the first time and assisted in improving public health.

As clinicians and practitioners, we are constantly using or generating big data, often without awareness. Every patient admitted, discharged, reviewed in clinic or collecting a prescription generates valuable data which can be harnessed for improvements in care. One benefit of big data in medicine is that it is from real people (patients) using the NHS rather than the rarefied populations that take part in clinical trials.

Often, the power of big data is harnessed by data linkage. Data linkage is the process of combining datasets to ensure information on an individual is linked to several different outcomes (e.g. medications prescribed, number of and reason for hospital admissions). The concept of data linkage is not a new phenomenon. Throughout the 20th century, data linkage (or record linkage) has become an important aspect of public health practice. In 1946, Halbert Dunn (Chief, US National Office of Vital Statistics) described linking birth and death records:

Each person in the world creates a Book of Life. This Book starts with birth and ends with death. Its pages are made up of the records of the principal events in life. Record linkage is the name given to the process of assembling the pages of this Book, into a volume.

Health informatics is a term that encompasses using data linkage to create large electronic datasets to improve the health and care of patients and the public.

Relevance to all – not just research

Big data are not only relevant to researchers. The implications for practising clinicians, patients and researchers of

¹Clinical Research Fellow, Geriatric Medicine, Alzheimer Scotland Dementia Research Centre and Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK and Specialist Registrar Geriatric and General Medicine, NHS Tayside; ²Clinical Research Fellow, Geriatric Medicine, Alzheimer Scotland Dementia Research Centre and Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK; ³Co-Director, Alzheimer Scotland Dementia Research Centre, University of Edinburgh, UK; Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK, and Centre for Dementia Prevention, University of Edinburgh

understanding healthcare, advice on healthcare policy and designing services is significant. In a digital world we are surrounded by data. Within medicine, data sometimes seem separate from our clinical job, and yet they are an integral part of delivering patient care and developing patient services. Since their introduction, electronic health records have collected vast amounts of information on the patient, the healthcare provider and the delivery of healthcare services. It is through these data that resources for clinical work can be allocated or a greater understanding of our patients' diseases or risk factors at a local, regional or national level can be gathered to improve patient care.

Everyone working in the NHS is by default involved with big data generation for current and future research and the wellbeing of patients, whether they are researchers or not. Therefore we all have a duty to understand how the recording of our patient-clinician interactions and diagnoses can influence the veracity of the big data that are later analysed, as well as how we can all realise the potential of big data in our individual specialities.

An example of big data in clinical practice is the Scottish Care Information Diabetes Collaboration (SCI-Diabetes), established in 2002.² Real-time clinical information is collected in a fully integrated and shared electronic patient record by clinicians. It crosses the primary-secondary care interface and tracks information on all 300,000 patients with diabetes in Scotland. The detailed data gathered provide valuable insights into the diabetes community, including trends in treatment and clinical outcomes.

The Farr institute has highlighted 100 ways of using data to make lives better in their case examples, many using Scottish data.³ A powerful example from Scotland highlights the key role clinicians have in ensuring the accuracy of current diagnoses and past medical history recorded on electronic health records.⁴ A team at the Universities of Stirling and Glasgow and NHS Information Services Division Edinburgh looked at suicides linked with hospital records. A quarter of suicides were people who died within three months of their last hospital (general and psychiatric) discharges. Only 14% discharged had a mental illness recorded during their last hospital visit despite one in five having a history of previous mental illness.^{4,5} This study highlights gaps in the current management and follow-up of these individuals and helps identify areas for targeted training and intervention to address this problem. These findings come from everyday practice but it is only by analysing these data can we find ways to prevent similar failures happening again. This highlights the crucial role all frontline staff have in contributing to the veracity of big health data.

The Twitter hashtag coined by the Farr institute is #DataSavesLives. This exemplifies the Farr Institute's aim of highlighting how data can save lives by using health data in research to give new understanding and insights, which benefits patients and improves public health.

Big data in Scotland

Procedures for storing and accessing big data vary internationally and within the four nations of the UK. Some countries legislate for the use of data (e.g. Care Act 2014, England and Wales) and linking data sets, and others have a more pragmatic approach (e.g. Scotland).

A key feature to facilitate the linking of big datasets is to have a universal identifier present across all datasets, known as a unique identifier. In Scotland, the Community Hospital Index (CHI) is a unique identifier which includes the person's date of birth (the first 6 digits) followed by a unique four digit ending (which indicates region of birth [or first registration] and gender). The decision in the 1970s for every person registered with a GP in Scotland to be allocated this unique identifying number from the CHI (centrally maintained) register was invaluable. The Scottish Record Linkage System, using the CHI, started in the late 1980s. Compared with the NHS number, the CHI number is a person-specific identifier and must be used in every hospital or GP episode. Although the NHS number is a unique identifier, it was not primarily used for linking data. As it does not contain the patient's date of birth it is harder to ensure matched records belong to the same person. The CHI system makes this process much more efficient as it makes access to routine clinical data and health episodes for research easier and less expensive.

One of the most striking recent examples of using big data in Scotland is the evaluation of the effects of the smoking legislation introduced in 2006.⁶ Researchers were able to evaluate trends in hospital admissions to determine the impact of this national policy on the health of Scotland's citizens.^{7,8} Without using routinely collected big data, this evidence would have been more difficult, and expensive, to collect, relying on bespoke data collection.

Through this unique identifier, Scotland has created the Scottish Longitudinal Study, which incorporates census data linked to health records and education data.⁹ This is one of the world's largest administrative datasets. It demonstrates the feasibility of linking multiple data (health, demographic and socio-economic). This work is only possible due to the establishment of key centres in Scotland to ensure data linkage conforms to the key principles and governance of data linkage. These centres have developed over time as the recognition of the power of data linkage in improving healthcare and research is growing (Box 1).

Characteristics of big data

The three Vs of big data first described in 2001 by an American IT firm¹⁰ are still used and relevant now:

1. *Volume* – the scale/size of the data
2. *Velocity* – the speed at which it is generated
3. *Variety* – the varying forms the data is available in – images, text

Box 1 Overview of recent key centres in the development of big data in Scotland

Scottish Informatics Programme

SHIP³¹ was initiated in 2009, and involving four universities and the Information Services Division (ISD) (part of NHS National Services Scotland), the principal source of statistics and healthcare services in Scotland. SHIP was the first organisation developing data linkage in Scotland to improve healthcare of patients and public. It is a Scotland-wide research platform for collating, managing, governing and analysing electronic patient records. This work has now moved to the Farr Institute.

Farr Institute of Health Informatics Research

Dundee University was the original centre that led to the creation of the Farr Institute³² in Scotland. Following expansion, the Farr Institute is a well-recognised UK-wide clinical research collaboration. Scotland now has seven (six universities and NHS Health Scotland) out of the 24 academic institutions and health partners across England, Scotland and Wales who are part of this UK-wide research collaboration. NHS Scotland is the principle health data custodian in Scotland.

Administrative Data Research Network

ADRN³³ has centres in all four nations of the UK. The Administrative Data Research Centre (ADRC) for Scotland is led by the University of Edinburgh and incorporates seven major Scottish research centres. Administrative data refers to education, housing, transport etc. and that collected in specific surveys of health or care behaviours/ attributes, such as the Scottish Health Survey, which have been linked to health data.³⁴

Scottish Informatics and Linkage Collaboration

SILC³⁵ is a collaboration of ADRN, The Farr Institute Scotland and The Urban Big Data Centre.

However, many have argued the recent addition of a fourth and fifth: *Veracity* (the accuracy of the data) and *Value* (allowing new insights and understanding).^{11,12}

In medicine, this fourth point is of critical importance. If the data we have to analyse do not have sufficient accuracy, then the output and relevance to clinical practice will be questionable. When considering how big data can shape what we do now in healthcare, and may influence future work, it is important to consider the five Vs of big data.

Challenges and complexities of big data

There are many challenges in using big data to inform our understanding of health. These are partly exemplified by the rapid growth of data sets and constant generation of healthcare data from multiple sources. The challenges can be grouped into three key areas:

1. *Data governance*, including storage, transferring, sharing and privacy

2. *Data generation*, including capturing, curating, updating and accuracy
3. *Data output*, including analysis, querying large datasets and generating meaningful and reliable outputs

Data governance

There are numerous processes in place to safeguard the public and patients' interests. Navigating the path of research ethics and the variety of permissions required is never easy.¹³ These ensure that non-identifiable data are used and that it is for a worthy cause in the public's interest.

Public concern exists over use of patient data and the potential for commercialisation.^{14,15} Public engagement has emphasised that researchers must ensure that health and social care data linkage research does not risk identifying individuals, is secure and that it demonstrates benefits for patients and other service users or the public.¹⁶ Therefore, it is appropriate that there are strict guidelines, codes of conduct and governance that surround accessing patient data in Scotland. It is the role of everyone in the NHS to understand the governance and strict protocols in place that govern data linkage and analysis.

Data generation

It takes time to generate large datasets. Rarely are they complete and establishing the best method of linking datasets, to ensure the person with missing data is correctly linked to the related records, is vital. Advances in methodological developments of data linkage are rapid.¹⁷

One key challenge for analysing big health data is the ability to ascertain diseases through the accuracy of clinical coding. Diagnoses recorded by clinicians are coded according to the ICD-10 code of diseases¹⁸ Researchers are mindful of ensuring these diagnoses are accurate when using big data sources,¹⁹ rather than individual patient assessment methods.

Data are only as good as what has been entered by the individual doctor or nurse and then translated or transcribed by the coders in medical records. This is even more important when looking at large volumes of data; the belief that larger numbers result in a more accurate picture is not necessarily true.

Data output

If all diagnoses are transcribed but the coding is incorrect, this could result in erroneous calculations of the prevalence of different diseases and, importantly, the reasons for admission to hospitals. This has far reaching consequences when we combine this incorrect information with the basis for funding services or driving forward health policy. We know from several articles that incorrect coding adds to this inadequate knowledge of people who use the NHS.²⁰ In times of a cash-strapped and over-stretched NHS, having accurate data on who uses the NHS and the conditions they present

Box 2 The Scottish Morbidity Record

SMR 01 data lists all admissions and day cases where an individual is admitted to an NHS general hospital in Scotland according to their ICD code.³⁶

What is recorded?

While there may be more than one diagnosis or medical problem during an admission episode, these may not all be included when translated from the paper or electronic records to the SMR 01 data. Additionally, if the coders in medical records are busy then only the first diagnosis is entered.

This is of relevance clinically, ensuring the first diagnosis is the most accurate in reflecting the patient's clinical picture. It is also important when considering national rates of diseases and reasons for hospital admissions.

National prevalence of diseases

Understanding the number of people in Scotland recorded with dementia is dependent on those with dementia being diagnosed and then clinicians recording dementia in clinical records. However, dementia is not commonly the principal reason for an acute admission to hospital, so it may not feature on the discharge summary unless recorded in 'other diagnoses' section. Therefore, a researcher looking for a disease such as dementia (rarely recorded as a primary reason for admission) may believe there are fewer people with a diagnosis of dementia using hospital admission data alone. We could draw the conclusion that those with dementia admitted to hospital, from the SMR 01 records, could be a far lower prevalence than would be accurate.

with is relevant to all of us. The Scottish Morbidity Record (SMR) in Box 2 illustrates how diagnoses are recorded and used for understanding diseases.

The accuracy of clinical coding must start from admission onwards. Discharge is an equally important time for ensuring accurate information to allow ICD codes for prevalence of diseases to be ascertained. All healthcare staff must be informed how the data they generate are used in order to help them understand the vital role of clinical coding and discharge correspondence to improve healthcare.

Social care data

Social care data can also provide a valuable opportunity to understand variations in patients' journeys. Thirty-two local authority areas in Scotland have responsibility for the provision of social care²¹ to the larger 14 regional health board areas.²² Therefore, a health board will include residents from multiple local authorities. This distinction can manifest as variation in patient services available within the same ward. This includes access to home care services and care home funding. As social care support directly influences the ability to safely discharge patients, understanding the variation across Scotland is vital, particularly as our population ages

and care needs increase. The Scottish Government has commissioned the Information Services Division to work with local authorities and NHS boards in the Health and Social Care Data Integration and Intelligence Project.²³ A key aim is to develop linked, shared databases containing individual-level anonymised health and social care data to evaluate interventions, policies and trends.²³ Central analysis will allow comparisons between local authority areas. By identifying variation, regions can consider each other's services and learn from examples of good practice to improve outcomes, in keeping with a key strand (reducing unnecessary variation in practice and outcomes) of the Chief Medical Officer's report.²⁴

The Scotland Social Care Survey is collecting information on home care usage, community alarms, telecare, and self-directed support.²⁵ The Care Home Census, collected since 2002, is another source of social care data in Scotland. It provides aggregate data on care homes, places, residents, length of stay and individual-level data on long-stay admissions.²⁶ Unlike the Social Care Survey (which produces a snapshot of use at one time), the annual Care Home Census aims to capture activity throughout the year; however, the completeness of these data has not been formally assessed. Both resources will have CHI added to help facilitate linkage with other routinely-collected data sources.

Analysis of English social care data suggests greater issues with data quality than those found in health data, although the authors propose some of this could be addressed by linking the two, maximising their usefulness.²⁷ A pilot project in NHS Tayside found that differences in documentation, recording and variables meant the work was feasible but challenging and required joint working to understand organisational culture and practice.²⁸ Building and maintaining cross-organisational relationships takes time.²⁸

The future

Scotland is poised to take advantage of the big data revolution in healthcare due to the unique identifier (CHI), one data custodian (NHS Scotland), and collaborative organisations for data linkage. Few countries have the ability to link high quality data of national coverage to allow patient-based analysis and follow up, and as the drive to integrate health and social care data increases this will provide a valuable additional resource to understand the patient journey.

Routinely collected health and social care data can be used to ensure practice is evidence-based, by examining in real-time the effects of changes at a local or national level on the outcomes of those directly affected. Comparison between areas to identify and explore variations allows targeted work to learn from and share good practice. Following the integration of health and social care in law,²⁹ the impetus is there to deliver improvements to the lives of the people of Scotland through more effective joint working.

There is tremendous potential to improve practice and conduct relevant research using routinely-collected data to

What can clinicians do to improve the quality of healthcare big data?

- Discharge letters – when listing several diagnoses linked with a patient's admission, consider the principal diagnosis that you want to be recorded as the first diagnosis. This may be the only diagnosis later used for establishing hospital admission data by disease within SMR data
- Death certificates – when issuing a death certificate, ensure that all relevant key medical diseases not linked with the primary cause of death are recorded in the secondary cause of death. This will allow more accurate national recording of prevalence of diseases at death
- CHI number – when documenting patient information, ensuring the CHI number (or appropriate unique identifier) is recorded on all patient records is key to facilitating record linkage. In many cases, where work is done electronically, this is automatically populated, but in situations where the clinician needs to complete records manually (e.g. handwritten prescriptions) it will allow a complete picture of the patient's journey to be recorded through data linkage

help understand how to work towards delivering optimal care at home or in a homely setting, the goal outlined in both the Scottish Government's Vision 2020 and the Six Essential Actions to Improving Unscheduled Care.³⁰

Funding

KEW and JKH are supported by clinical research fellowships from Alzheimer Scotland and the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/L501530/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. The funders played no part in the design or conduct of this paper. ①

References

- 1 What's a Byte? Megabytes, Gigabytes, Terabytes What are they? <http://www.whatsabyte.com> (accessed 18/1/17).
- 2 Scottish Care Information Diabetes Collaboration. <http://www.sci-diabetes.scot.nhs.uk> (accessed 18/1/17).
- 3 The Farr Institute of Health Informatics Research. 100 Ways of Using Data to Make Lives Better. <http://www.farrinstitute.org/public-engagement-involvement/100-ways-of-using-data-to-make-lives-better> (accessed 18/1/17).
- 4 University of Glasgow/University of Stirling/NHS National Services Scotland/ISD Scotland. Can Data Be Used to Help Prevent Suicides? <http://www.farrinstitute.org/wp-content/uploads/2016/10/11.-Can-Data-Be-Used-To-Help-Prevent-Suicides-print-1.pdf> (accessed 18/1/17).
- 5 Dougall N, Lambert P, Maxwell M et al. Deaths by suicide and their relationship with general and psychiatric hospital discharge: 30-year record linkage study. *Br J Psychiatry* 2014; 204: 267–73.
- 6 Haw SJ, Gruer L, Amos A, Currie C et al. Legislation on smoking in enclosed public places in Scotland: how will we evaluate the impact? *J Public Health* 2006; 28: 24–30.
- 7 Pell JP, Haw S, Cobbe S et al. Smoke-free legislation and hospitalizations for acute coronary syndrome. *N Engl J Med* 2008; 359: 482–91.
- 8 Mackay D, Haw S, Ayres JG et al. Smoke-free legislation and hospitalizations for childhood asthma. *N Engl J Med* 2010; 363: 1139–45.
- 9 Introduction to the Scottish Longitudinal Study. <http://sls.lscs.ac.uk/about> (accessed 18/1/17).
- 10 Laney D. 3D Data Management: Controlling Data Volume, Velocity and Variety. 2001. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed 18/1/17).
- 11 IBM. Infographics & Animations: The Four V's of Big Data. <http://www.ibmbigdatahub.com/infographic/four-vs-big-data> (accessed 18/1/17).
- 12 Burbank D. The 5 V's of Big Data. *Enterprise Architects*. 2016. <http://enterprisearchitects.com/the-5v-s-of-big-data> (accessed 7/2/17).
- 13 Brett CE, Deary IJ. Realising health data linkage from a researcher's perspective: following up the 6-Day Sample of the Scottish Mental Survey 1947. *Longitudinal and Life Course Studies* 2014; 5: 283–98.
- 14 National Data Guardian. National Data Guardian for Health and Care: Review of Data Security, Consent and Opt-Outs 2016. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF (accessed 18/1/17).
- 15 Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. *J Med Ethics* 2015; 41: 404–9.
- 16 Scottish Government. Linking Social Care, Housing and Health Data: Social Care Clients' and Patients' Views. 2011. <http://www.gov.scot/Publications/2011/09/20085846/11> (accessed 8/1/16).
- 17 Harron K, Goldstein H, Dibben C, editors. *Methodological Developments in Data Linkage*. Chichester: John Wiley & Sons, Ltd; 2016.
- 18 International Statistical Classification of Diseases and Related Health Problems. 10th Revision. 2016. <http://apps.who.int/classifications/icd10/browse/2016/en> (accessed 18/1/17).
- 19 Woodfield R, Grant I, Sudlow CL. Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from the UK Biobank Stroke Outcomes Group. *PLoS One* 2015; 10: e0140533.
- 20 Medford AR. Aberrant clinical coding: an underappreciated entity for NHS Trusts, clinicians and coders. *J R Coll Physicians Edinb* 2013; 43: 101–2.
- 21 Scottish Government. Scottish local authorities 2016. <http://www.gov.scot/Topics/Government/local-government/localgov/usefullinks> (accessed 5/6/17).
- 22 Scottish Government. NHS Boards. <http://www.gov.scot/Topics/Health/NHS-Workforce/NHS-Boards> (accessed 22/12/16).
- 23 NHS National Services Scotland/ISD Scotland. A Guide to Data to Support Health & Social Care Partnerships in Joint Strategic Commissioning and Joint Strategic Needs Assessment. 2014. <http://www.isdscotland.org/Health-Topics/Health-and-Social-Community-Care/docs/Guide-to-Data-to-Support-HSCPs-8Oct2014-FINAL.pdf> (accessed 22/12/16).

- 24 NHS Scotland. *Chief Medical Officer's Annual Report 2014–15: Realistic Medicine*. <http://www.gov.scot/Resource/0049/00492520.pdf> (accessed 28/1/16).
- 25 Scottish Government. *Social Care Services, Scotland, 2016*. <http://www.gov.scot/Resource/0051/00510699.pdf> (accessed 22/12/16).
- 26 Information Services Division Scotland. *Care Home Census. Scottish Statistics on Adults Resident in Care Homes*. 2015. <http://www.isdscotland.org/Health-Topics/Health-and-Social-Community-Care/Publications/2015-10-20/2015-10-20-CHCensus-Report.pdf?> (accessed 20/10/15).
- 27 Kemm JR, Robinson J, Verne J. Social care data in England: what they tell us and what they do not tell us. *Public Health* 2010; 124: 265–8. <http://www.sciencedirect.com/science/article/pii/S003335061000065X>
- 28 Witham MD, Frost H, McMurdo M et al. Construction of a linked health and social care database resource – lessons on process, content and culture. *Inform Health Soc Care* 2015; 40: 229–39.
- 29 Scottish Government. *Integration of Health and Social Care*. 2016. <http://www.gov.scot/Topics/Health/Policy/Adult-Health-SocialCare-Integration>
- 30 Scottish Government. *6 Essential Actions to Improving Unscheduled Care*. <http://www.gov.scot/Topics/Health/Quality-Improvement-Performance/UnscheduledCare/6-Essential-Actions-To-Improving-Unscheduled-Care> (accessed 16/1/17).
- 31 ScottishH Informatics Programme (SHIP). <http://www.scot-ship.ac.uk/index.html> (accessed 18/1/17).
- 32 The Farr Institute of Health Informatics Research. <http://www.farrinstitute.org> (accessed 18/1/17).
- 33 Administrative Data Research Centre Scotland. <https://adrc.ac.uk/about/network/scotland> (accessed 18/1/17).
- 34 Hanlon P, Lawder R, Elders A et al. An analysis of the link between behavioural, biological and social risk factors and subsequent hospital admission in Scotland. *J Public Health* 2007; 29: 405–12.
- 35 Scottish Informatics and Linkage Collaboration. <http://www.datalinkagescotland.co.uk> (accessed 18/1/17).
- 36 ISD Scotland Data Dictionary. SMR01. <http://www.ndc.scot.nhs.uk/Dictionary-A-Z/Definitions/index.asp?Search=S&ID=997&Title=SMR01> (accessed 18/1/17).

The College Journal Prize



ROYAL
COLLEGE of
PHYSICIANS of
EDINBURGH

The College Journal Prize 2016, sponsored by the Senior Fellows' Club, was won by RT Tran and AJ Rankin (joint first authors) et al. for their paper 'Short runs of atrial arrhythmia and stroke risk: a European-wide online survey among stroke physicians and cardiologists'. This paper can be read in issue 2, 2016, http://www.rcpe.ac.uk/sites/default/files/jrcpe_46_2_tran.pdf

A prize of £250 will be awarded to the first-named (or corresponding) author of an original research paper on a clinical topic, deemed by a panel of judges to be the best paper by a doctor-in-training (i.e. pre-consultant level) published in the *Journal of the Royal College of Physicians of Edinburgh* in issue 4, 2016 and issues 1, 2 and 3, 2017. The prize-winner will be invited to give a short oral presentation based on their paper at the Medical Trainees' Conference 2018.

Further details may be obtained from the Editorial Office, Royal College of Physicians of Edinburgh, 9 Queen Street, Edinburgh EH2 1JQ

Tel +44 (0)131 247 3666 or email editorial@rcpe.ac.uk

This final published PDF is included in line with the journal re-use guidelines for Green Open Access for using the final published PDF as author *"in your dissertation or thesis, including where the dissertation or thesis will be posted in any electronic Institutional Repository or database"* with author contribution and appropriately referenced. Author contributions are outlined in the Acknowledgements section of this thesis.

In summary, the aim of this publication was threefold: 1) to highlight the current healthcare data landscape in Scotland at the time, 2) to highlight how clinically collected routine digital data can be used for research, and 3) to foster a connection between the clinician generating the data and the researcher later using it. Data are only as good as the data entered into the dataset, e.g. the wrong, or missing clinical information are subject to human error in entering the data and may have implications in how conclusions are drawn from data analysis by researchers. The generation of clinically accurate data in busy clinical scenarios, with systems to ensure accurate coding, is important. Data in datasets for researchers to analyse is affected by: the validity of the data entered manually into the original dataset, the relevance of what the data captures (or whom it captures), and where it captures (e.g. in healthcare does it capture a set part of a patient's journey or does it cross different community/secondary care settings).

A theme of this paper is highlighting how the quality of routinely collected health data can impact on later using the data secondarily for research. This is an important aspect and I discuss this in this chapter. Linked, but separately, is the issue of transcribing this information from a clinical encounter to the databases.

Whilst much of this paper still stands true for big data in healthcare in Scotland today, there have been a number of changes in the processes and overarching governing bodies which are protecting and authorising access to its secondary use for research. I will therefore discuss these changes in Sections [3.10.1](#) ([UK health](#)

[data landscape](#)) and [3.10.2 \(Current Scottish data strategy\)](#) as part of the current landscape of health data research in the UK, including variation across the four nations of the UK.

In this chapter I address some key areas of using routinely collected health data for research. These include types of data, data governance, access and permissions in general, data linkage and the framework in which health data research should take place. Following this, I will focus on how routinely collected health data can be used for dementia research as this is important for my research question and subsequent chapters (Chapters [4,5,6](#)) as I use these approaches.

3.3 Considerations of using routinely collected data for research

Today, the vast amounts of routinely collected health data form a valuable (secondary re-purposed) resource. This can increase our understanding of both our patients and variations in health and disease in real time, in addition to the reason they were collected, to deliver healthcare. Using routinely collected health data for research is complementary to traditional methods of research (such as primary data collection), rather than in place of them. One benefit of big data are that it can include all patients using healthcare systems. It does not only include very selected populations of patients classically used in clinical trials. Big data does not exclude older frail patients, including those with dementia, who may not easily be able to take part in traditional research methods.

There are some excellent examples of using routinely collected health data for research purposes. In 2007 in the UK, the first to develop this resource as a secondary use for research, or as they described it “*unlock data from the full electronic medical record*”, was the Clinical Record Interactive System (CRIS) in South London and Maudsley NHS Foundation Trust [[439](#)]. This was a collaboration that developed between the National Institute for Health Research (NIHR)

Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London.

Importantly, patients were a key part of the development of this collaboration. As South London and Maudsley NHS Foundation Trust was Europe's largest mental health provider, this was a powerful collaboration to advance the use of routinely collected healthcare data (electronic health records) for research uses. It allowed researchers to access a large volume of routinely collected data. There have been a large number of publications using this collaboration with some focusing on how diseases develop or identifying people at risk or which treatments are most cost-effective, with the aim to improve patients' care and outcomes [440, 441, 442, 443, 444, 445]. There are also several publications using this collaboration investigating dementia research.

One of the strengths of this collaboration was that patients were involved directly in the start of the process from design, implementation as well as ongoing input and direction. The patient feedback has been positive. It is also 'opt out' and is therefore more representative of everyone using the services. In summary, not only was this a first in the UK but it importantly involved patients which, as it is *their* routinely collected health data, should be the gold standard, wherever possible. It also involved lived-experience research ambassadors giving service users the opportunity to be directly involved in the CRIS research, creating an inclusive health data research approach.

The public's opinion in how their data are stored, linked and re-purposed for research is important. Aotearoa New Zealand commissioned a report discussing public attitudes to data integration when creating their national routinely collected data repository, The Integrated Data Infrastructure (IDI) [446]. The IDI is discussed in detail in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#). However, there can still remain concerns about linked data from public groups [447].

Patient and public involvement and engagement (PPIE) is vital for healthcare research. This is extremely important also when re-purposing routinely collected health data for research [448, 449]. In PPIE researchers work with patients (and/or patient advocacy groups) and the public to shape and design research. The National Institute of Health and Care Research (NIHR) UK has clear guidance and recommendations for researchers regarding involving and engaging with patients and the public for research [450].

The motto for PPIE *"nothing about us, without us"* still stands for research using routinely collected data [451]. In many ways extra work is required for this type of PPIE to overcome past mistrust of the proposed use of routinely collected data for research. In Scotland (as highlighted more in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#)), the Health and Social Care Public Benefit and Privacy Panel (HSC-PBPP) has an important role in ensuring the research is in the public's interest [452].

There are UK standards for public involvement in research as part of a UK-wide partnership of organisations including NIHR, Chief Scientist Office, Health and Social Care Public Health Agency, and Health and Care Research Wales [453]. The six UK standards are: inclusive opportunities, working together, support and learning, governance, communications, and impact [453], whose aims are to be: *"a framework for what good public involvement in research looks like and are adaptable to different situations; designed to encourage reflection and learning, including where lessons have been learned when public involvement has failed to lead to expected outcomes; a tool to help people and organisations identify what they are doing well, and what needs improving; and, intended to be used with any method or approach to public involvement in research"* [453].

Within dementia research, work has been undertaken to embed patient and public involvement in dementia research through reflections of experiences in a ran-

domised controlled trial [454]. A number of groups including the Scottish Dementia Working Group Research Sub-Group investigating core principles for involving people with dementia in research [455], and other groups have investigated how to effectively include people living with dementia in research [456, 457, 458].

In big data healthcare research using routinely collected data there are recent scoping reviews investigating the use of PPIE for this type of research [459, 460, 461]. Many of the reviews prior to these have investigated the trust, or mistrust, from the public of using routinely collected data/big data for research [462, 449, 463, 464].

In recent years, the failure to include and consider the views of those whose data are being re-purposed (even if in an anonymised and de-identifiable format) has led to mistrust and barriers in collaborative approaches. In my published review in Section 3.2 (Peer-reviewed published manuscript) [2], I highlighted one such example, the UK *care.data scheme* [465] as an example of loss of trust from the public in re-purposing of routinely collected health data for research. PPIE in routinely collected data for health and social care research is a growing area. There are many considerations, but it is important that PPIE is not perceived as a tick-box exercise or tokenism as has been highlighted [466]. Additionally, there is growing awareness in healthcare research in general of the need to fully describe PPIE in published research outputs using reporting tools such as the Guidance for Reporting Involvement of Patients and Public version 2 (GRIPP2) [467].

Ensuring key stakeholders are involved in the re-purposing of routinely collected health and social care data for research in the early stages is important. Additionally, there are also further considerations of using this type of data for research including the variability and reliability of the data between different systems.

3.4 Healthcare data: types and coding

Each country collects data in different ways, including across the four nations of the UK. I discuss the current landscape within the UK in detail in Section 3.10.1 (UK health data landscape). Data found within routinely collected health data which can be valuable for research includes numerous variables including age, sex and demographics of patients; diagnoses, investigations undertaken, treatment initiated or stopped, medications, inpatient or outpatient care, type of department (e.g., cardiology, geriatric medicine, surgery etc.).

Aside from health data in routinely collected health data there are also administrative health data generated through an individual interacting with healthcare. Social care data are additional important information. Administrative data in the NHS refers to information collected for the operational purposes rather than the healthcare interaction related to health or disease investigations or management. For instance, in the UK the NHS collects information on admissions and appointments to monitor trends in hospital activity. Births and deaths may occur within a healthcare setting, and healthcare professionals can be involved in the generation of the data, but they are also included as ‘administrative’ data. Information on births and deaths are also collected by government bodies for monitoring.

There are multitudes of other routinely collected data that may be important for understanding a person’s health, well-being and diseases. Aside from private companies collecting data routinely on the population for numerous reasons (supermarkets, phone companies, smart phone apps) there are other types of routinely collected data that can be relevant for healthcare research. Additional administrative data (outside of the health setting such as Education, weather, pensions or population data) can be valuable, as well as social care data, which may be collected in different systems to which healthcare data is routinely collected.

The World Health Organisation (WHO) constitution states health is a: *“state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”* [468]. This is important to consider for person’s health as more than the sum of their diseases. Some countries (particularly Scandinavian countries) are already able to link all administrative, health and social care data [469, 470]. There is a growing drive to integrate health, administrative and social care data. All are important in understanding an individual’s health, well-being, interaction with society and how diseases impact on these aspects. However, there can be challenges in linking these data. Therefore, for the purpose of this section, routinely collected data refers to routinely collected health data.

How a diagnosis is recorded in routinely collected health datasets can be in various forms. A full diagnosis is classically shortened to a ‘code’ which represents that diagnosis. The code is normally the code recognised with the international classification of the disease such as the WHO International Classification of Diseases (ICD) coding system [21]. The process of this translation from a clinician writing a diagnosis on paper or electronic notes in free hand and it being converted to a recognised code is called ‘coding’ a diagnosis.

Not all routinely collected healthcare datasets use the same methods for coding diagnoses. This has traditionally produced a lot of challenges. In the UK, primary care and secondary care had different methods for ‘coding’ diagnoses. So, when trying to create that ‘book of life’ of a patient’s diagnoses, and their patient journey, in healthcare it was difficult because in essence the two different routinely collected healthcare datasets could not easily ‘talk to one another’ as they used different systems and importantly different coding methods. Primary care previously used READ codes and are now using Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [471].

Traditionally, secondary care (hospital) interactions used International Classifi-

cation of Diseases ICD codes (currently ICD version 11 is available, but not yet used across the NHS as many still using ICD version 10) [21, 145], and prescribing datasets use codes for drugs [472]. Previously, the British National Formulary, BNF, used a hierarchy to assign codes to drugs and chemicals and this was a specific way to show what was prescribed including the formulation [472]. This has now been replaced in NHS England with NHS Business Services Authority using a legacy version of the BNF hierarchy to assign codes to drugs and chemicals and map to SNOMED CT [473]. Scotland is beginning the process of mapping to SNOMED CT as I outline in Section 3.10.2 (Current Scottish data strategy).

Whilst the ICD coding system is international[21, 21], there have been changes with time (i.e., diseases that were recorded in ICD version 9 may have changed by ICD version 10 and then ICD version 11) and also between countries (i.e., some countries such as Australia and New Zealand use a modified version, ICD-AM, Australasian modification, with some subtle differences) [144, 474]. I discuss this further in Chapter 4: Investigating dementia in Aotearoa New Zealand (ANZ) using routinely collected data in relation to dementia codes in the ICD-10 AM, but also differences in how frequently they are updated with revisions.

Differences are important in the transition from different versions of ICD coding if there has been a change in how diseases are described. When ICD-9 changed to ICD-10 it was noted that there was increased sensitivity for dementia in the ICD-10 version which could potentially show an increase in prevalence if the two versions were compared directly [144]. These challenges could be expected for each version change of ICD. It is also important to understand for hospital inpatient stays that discharge letters will include a total of up to six ICD codes in Scotland (one main diagnosis and five ‘other diagnoses’) and that this will be decided not only by the clinician creating the discharge summary but also those working in medical records with coding expertise.

The drive now internationally is to use Systematized Nomenclature of Medicine

Clinical Terms (SNOMED CT) [475]. This is a structured clinical vocabulary for use in an Electronic Health Record (EHR) [475]. It is found to be the most comprehensive and precise clinical health terminology and can allow communication between different clinical health records. In NHS England, SNOMED CT has been created using the international edition of SNOMED CT and additional UK-specific content with the UK clinical extension, UK drug extension, and the UK pathology extension mapped to ICD-10 and the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures 4th revision (OPCS-4) [471]. In Scotland, Wales and Northern Ireland there are programmes underway to follow this method.

As highlighted in 2023 at the World Health Assembly with statements from a number of countries (including the UK and New Zealand), there are international calls for the World Health Organisation (who produce the ICD versions) to do more for harmonisation between ICD-11 and SNOMED CT now there is an international drive to use SNOMED CT [476].

There is also much more information which is captured in routinely collected data as ‘free text’ that may not be recorded with a clear code to extract. This is typically information on interactions, notes between clinician and patient including symptom details, or within imaging reports. There is ongoing research using a variety of methods, including AI, to ‘mine’ this valuable data resource which may often have important information that does not reach the discharge letter or coded entry in the dataset [477].

3.5 Data linkage: unique identifiers, types of linkage, datasets

Data linkage is the ability to link one dataset (e.g., patients’ hospital discharges and medical conditions) to another (e.g., mortality data and cause of death) [478]. Much emphasis has been placed on data linkage being a new phenomenon. It is

not a new concept. In the 1940s, the Chief of the US National Office for Vital Statistics (Halbert Dunn) described linking birth and death records for a person as their ‘Book of Life’ [479]. Through data linkage of data collected routinely for clinical care, the patient’s ‘Book of Life’ can be a unique resource for answering research questions, whilst also being much more cost efficient than traditional research methods in a resource limited setting.

There are differences in the ability to link data between countries and within countries. Some of that stems from data governance, access, permissions and the legal framework and some of that from the actual mechanisms to be able to link different healthcare datasets. Both of these factors can influence the timely access for the researcher to be able to undertake research using routinely collected healthcare data. I will now discuss some key aspects of data linkage: what is meant by a unique identifier, the types of linkage and linking external datasets to routinely collected health data.

3.5.1 Unique identifiers

Unique identifiers, in the context of routinely collected health data research, are unique universal numbers assigned to an individual that transfer across different healthcare settings [478]. This means that any interaction with healthcare services can be found for an individual by searching for their unique number (unique identifier). Practically speaking within a clinical context, it means an individual’s lists of investigations, diagnoses, management etc can be found (when stored digitally) by using that unique number to search within the clinical datasets. More than one person can have the same name or even the same name and date of birth. Technically, no one should have the same unique identifier in healthcare and it should stay with you throughout your life. These unique numbers are often referred to as a Unique Patient/Person Identifier(UPI), a Unique Identifier(UI), a Universal Health Identifier(UHI) or a personal identity number.

When using routinely collected health data for research then the ability to link

data at a large scale in healthcare rapidly is made easier by the use of a unique patient/person identifier (UPI) [478]. This is because a UPI should be universal across healthcare settings, therefore transferable across different datasets [478]. Within the UK, data linkage cannot be quickly and easily undertaken across all four nations because individuals do not have UPIs that span all four nations of the UK. The NHS system is separated across the four nations. There are numerous developments to try and address this complex problem because of the lack of a UK-wide UPI.

On the other hand, in some countries, a UPI is given to an individual at birth and is not purely for healthcare and is assigned by the Tax Office and used in many different contexts (administrative, social and health) allowing linkage of different types of data [4, 480, 470, 481]. A UPI should stay with you throughout your life in that country so new UPIs should not be generated and create confusion. The benefits of a UPI are that it creates a ‘spine’ within datasets where multiple datasets can be linked and ‘hung’ off the ‘spine.’ This spine of UPI ensures that each individual has one UPI entry which can link multiple hospital discharges, prescription admissions, healthcare interactions together into one large ‘big data’ set.

In the UK, everyone within the National Health Service (NHS) has an NHS number. Whilst this is a unique number, there are complexities in the NHS number. Technically, an individual could be given more than one NHS number if they are seen in different NHS health boards and the prior NHS number is not immediately available so a new one is generated. It is easy to link multiple NHS numbers with the one individual if the other personal identifiable information is unique. However, common names i.e., John Smith and the same date of birth means carefully linking the ‘correct’ John Smith’s multiple NHS numbers can be challenging.

In Scotland, all individuals interacting with healthcare have a UPI. This is used

instead of the NHS number (which individuals also have) as the primary method for searching for patients. Scotland's UPI in healthcare is a 10-digit unique number called the Community Health Index (CHI) [482]. This was established in 1980s and every person living in Scotland (or any visitor using the healthcare system) is given this UPI [482]. It includes their date of birth plus two digits indicating their region of birth (or first registration), a digit indicating sex, and a test digit (e.g., 0101201175) [482]. This is used in place of the NHS number for patient interactions with the Scottish healthcare system [482]. From a clinical perspective it can create a short delay, whilst a CHI number is generated, if someone is admitted acutely who is not resident in Scotland and investigations are required. As the CHI number is a highly identifiable number (as it includes a person's date of birth) there is a CHI Linkage and Indexing team within Public Health Scotland that subsequently create a secure process to allow data linkage of an individual's CHI number for research[483]. This process involves both CHI seeding and then CHI indexing[483].

CHI seeding is defined as: *"a process where a person's current CHI number can be added to a dataset using probabilistic record linkage methodology. Records are matched against our CHI downloads using demographic variables such as: fore-name(s), surname, date of birth, sex, postcode of residence"* [483]. CHI indexing is defined as: *"CHI indexing is a process in which CHI numbers, which are highly identifiable, are replaced with a unique key (index number)"* [483]. I discuss the the additional changes planned in Scotland in Section 3.10.2 ([Current Scottish data strategy](#)). Aotearoa New Zealand equally has a UPI which will be discussed in more detail later in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#).

Historically, for the rest of the UK, UPI use has varied over the years. Regions in England have long established UPI research databanks of patients. This includes the Clinical Practice Research Datalink - a network of primary care practices with linkage to a range of datasets including hospital datasets [484]. The CRIS

database in South Maudsley NHS Trust mentioned in Section 3.3 ([Considerations of using routinely collected data for research](#)) is another example [439].

In NHS healthcare in England, there are data collected on hospital discharges across NHS England via the Hospital Episode Statistics (HES), and held by NHS Digital in England [485]. NHS England has a national Secure Data Environment (SDE) [486]. In 2004, NHS England Digital created a ‘Spine’ for England, Wales and the Isle of Man, which moved to version two in 2014 [487]. The Spine includes a UPI and allows information to be shared securely at a national (England, Wales and Isle of Man) level using datasets such as the Electronic Prescription Service (EPS), Personal Demographics Service (PDS), National Care Records Service (NCRS), e-Referral Service and the Patient Flags Service (API) [487]. It joins the Information Technology (IT) infrastructure for health and social care across these areas of the UK and it is being used daily by healthcare professionals [487].

Additionally, a recent report (2022) by the Health Foundation in England outlines the value of harnessing the potential of routinely collected data [167]. OpenSAFELY is a collaborative partnership in England providing a secure analytics electronic platform for NHS electronic health records for researchers [488].

In Wales, an additional process has been established with a UPI created within the Secure Anonymised Information Linkage (SAIL) databank [489]. In 2006, the Health Informatics Research Unit (HIRU) was created with funding from Health and Care Research Wales [490]. In 2007, a pilot study called the SAIL Databank was created initially starting around Swansea and then nationally rolled out across Wales with online access for researchers since 2011 [490]. The SAIL Databank consists of real-world anonymised individual-level population-scale data. there is a large repository of health, social and administrative data held within SAIL with the potential to explore health data research questions across the population of Wales as individuals have a unique identifier [490].

In Northern Ireland, data is stored within a safe haven called the Honest Broker Service [491]. This safe haven brings together routine data collected from health and social Care (HSC) services in Northern Ireland [491].

3.5.2 Types of linkage

Data linkage is undertaken through two common methods of data linkage: deterministic or probabilistic [478]. This can be complex and require experienced data scientists processing and linking data prior to a researcher using the data for their research question.

The first (deterministic linkage) requires a set of predefined rules and often involves a unique patient identifier as discussed above which can become a ‘spine’ [478]. For instance, to join two, and ultimately many more, datasets to give a more complete picture of the patient’s journey, there needs to be a unique ‘spine’ of the dataset which remains constant and is present in all datasets [478]. If datasets requiring linkage have a UPI, then an appropriately trained researcher can easily link these (once appropriate access, permissions and framework have been approved) by using statistical programming software such as R, SQL, STATA, SAS, Python.

The second method, probabilistic data linkage, determines the likelihood that records in different datasets match the same person by using statistical weights and thresholds using various non-unique identifiers such as name, date of birth, address [478].

In reality, outside countries that use a UPI from birth for all interactions with healthcare, administrative and social care data, for most data linkage for healthcare research there is a mixture of datasets with a UPI that can be deterministically linked and others that have different identifiers that require probabilistic linkage with expert data scientist involvement. In this context, datasets are normally referred to as ‘nodes’ with multiple ‘nodes’ linked to a ‘spine.’ For clinical

research, the term datasets instead of nodes are more commonly used.

The gold standard in health data research is to have health (primary and secondary care), social care, and other administrative data linked. However, whilst healthcare data collection and linkage has been long established in many countries worldwide [4, 489, 470, 469, 492, 493, 2, 494, 495, 481, 496], social care data is far less established and there are still significant hurdles to achieve full health and social care big data linkage [2, 497]. Within the UK, there are variations in how health data are linked. Across the UK, there are well established systems for linking health data [489, 2, 489] at an individual-level, but population scale, governed by appropriate permissions and procedures for researchers to access it [498]. Whilst this linkage is easier with a UPI, not all countries, even with a UPI, have the ability to easily link primary and secondary healthcare data. There are advancements in the linkage of routinely collected data in primary care [484], and separately amongst secondary care datasets, but integrating both together routinely remains a hurdle in many parts of the UK. This hurdle, along with the hurdles of incorporating social care data linkage, are important when we want a holistic view of a patient's journey.

3.5.3 Datasets: linking to routinely collected data

As demonstrated, routinely collected health data can be a powerful resource for researchers to use. However, it does involve an individual interacting with healthcare. Some in society may have limited interactions with healthcare, for various reasons, or be hard to reach populations through more traditional research methods [499]. They may be part of research groups or be part of birth cohort studies. When research is undertaken face-to-face to collect information specific to a certain research question then the data obtained is generated after careful consideration of what is needed to be collected in the initial stages. The challenge with routinely collected healthcare data is that it is not primarily developed for research, but for routine patient care. The types of information collected is not designed for research, although it still has valuable information that can be used

for research.

In essence, routinely collected health data gives a breadth of information on the population or population groups engaged with healthcare services, but without the depth that is traditionally required in research. One way to increase the information, and depth of information, for individuals within routinely collected health data is by linking external datasets to routinely collected health data [179, 181, 478, 180, 500, 501]. I discuss the processes of governance, access, permissions and safe environments for linking data in Section 3.6 ([Data governance, access and permissions](#)) as this is critically important to ensure both anonymity and that no one is indirectly identifiable. Research datasets (such as cohort studies) or other clinical datasets that can be linked and therefore ‘embedded’ or ‘nested’ with routinely collected health data provide a unique and valuable research resource.

Cohort studies can provide a valuable resource to investigate the role of different risks or protective factors for a variety of outcomes. Some cohort studies have valuable data from birth onwards allowing a life course perspective as the cohort ages. Some cohort studies have been created retrospectively such as the Lothian Birth Cohort [502] or Nun study [190] as the serendipitous discovery of early life data allowed a cohort study to be constructed. Routinely collected health and administrative data are often generated daily and, as discussed above, provide a valuable resource to investigate diseases over time. It is costly to continue frequent ‘waves’ of cohort assessments.

The information within cohort studies can be powerful with a depth that routinely collected data does not have. Linking a cohort within routinely collected health, and administrative, data provides a unique longitudinal resource, and life course approach, to investigate diseases. The potential to link cohort studies with routinely collected data allows 1) a continual ‘story’ of the person’s health and disease 2) a breadth of data by capturing other people within geographical areas that were not part of the cohort and 3) allows a researcher to adjust for

key confounders when investigating a specific outcome in routinely collected data that would be difficult to do as recording of a specific variable important for the research question may or may not be present in routinely collected health data.

As dementia prevalence increases with age, longitudinal studies are important for understanding life course risk, or protective factors, for dementia. Longitudinal cohort studies are important for this. However, routinely collected health data have also proved useful in allowing research questions investigating dementia epidemiology at large country level scales to be undertaken. Combining both cohort studies, or longitudinal studies, with routinely collected healthcare data for dementia research opens up an expanded data set for researchers to understand this disease, since as people age they may not be able to take part in traditional research studies in the same way.

The Dementias Platform UK (DPUK) Data Portal is a valuable data repository facilitating access to a large number of cohorts relevant to dementia within the UK secure eResearch Platform (UKSeRP) [503]. Many of these are also linked, or have the ability to be linked, to routinely collected health data such as Electronic Health Records(EHRs). As a recent paper highlights *“the legal status of accessing EHR data linked to research data using third party platforms is a grey area in the UK”* [503].

3.5.4 Data sharing

There is a careful balance between ensuring data sharing occurs to advance health-care research, avoids duplication, and improve the lives of individuals by greater understanding of health and disease. Data sharing must do this whilst also ensuring the protection of sensitive research data, the individual’s personal privacy and following the law that governs personal data.

To navigate this complexity, researchers are expected to generate metadata and data sharing agreements. Whilst there is ambiguity in the definition of what

‘metadata’ is, it could be considered: data (written descriptive information) about data (source data of interest) [504]. In health data research it is a way to list information about the data you are working on. There is also scope to list a contact should other researchers want to connect and collaborate. ‘Open access’ data refers to the unrestricted use of data and available for anyone to download. When data contains personal or sensitive research data (even if pseudo-anonymised), which is common in health data research, then ‘open access’ is not appropriate, or legally allowed. Within the metadata, a researcher can clearly highlight how collaborations can proceed for this type of data. Normally institutions have data sharing agreements with their institutions outlining the legal framework and process for sharing data.

Therefore, for some time (pre COVID-19) there was a growing drive within health data research in general to improve collaboration, data sharing and widening accessibility to datasets for researchers. Reducing repetition of research, increasing transparency of health research and ensuring that patient data was used in the most efficient way to improve understanding of health and disease process. The concept that developed from this for research was the FAIR Data Principles [505]. The acronym FAIR stands for Findable, Accessible, Interoperable and Reusable [505]. This became a growing requirement in data-driven research with justification required from researchers if they were not following this concept.

There was still some disconnect between clinicians creating the data found in routinely collected healthcare data and researchers using it secondarily for research. When the COVID-19 pandemic hit at the start of 2020 these requirements for data to be research ready and for data sharing to be a priority to understand and treat this deadly pandemic between clinicians, researchers and national government bodies merged rapidly. COVID-19 highlighted and accelerated the need to share data at a national and international level and one example in the UK was linking data across multiple sites in NHS England [506].

Since the COVID-19 pandemic, the World Health Organisation has also driven forward the need for data sharing, including uniformity in the ability to share data between countries. One important aspect that the WHO highlights is the need for “*leveraging data rather than simply collecting it*” [507]. Whilst this is essential during a pandemic situation, there is much to consider in relation to multiple long-term conditions and chronic diseases particularly dementia. As outlined earlier in this chapter, there is progress within dementia big data research of sharing cohorts with other researchers through the UK Dementias Platform [503].

3.6 Data governance, access and permissions

Strict processes, including a legal framework, are in place for a researcher to use personal clinical routinely collected health data for research. The data governance, access, and permissions processes vary between countries and even within countries to access different datasets. This variation exists for a number of reasons including different legal frameworks, public engagement and the content of the datasets. Some datasets may be ‘open access’ and therefore not confined to the same governance, access and permissions requirements. To understand why data governance, access and permissions are important for using routinely collected health data for research, it is important to understand two key factors first: the difference between pseudo-anonymisation and anonymisation. (cite)

Anonymisation refers to the process of personal data being irreversibly prevent re-identification leading to anonymous or non-personal data. Within GDPR, data is considered anonymous if there is *reasonable* likelihood that it cannot be linked to an identified or identifiable person. The data is then aggregated to the point where specific events can not be linked to individuals. One example of this are country statistic reports of the numbers of individuals with different diseases as a population level.

In comparison, pseudo-anonymisation refers to personal data being de-identified

and enables the personal data to become unidentifiable unless further information is available, or included. it is not irreversibly preventing re-identification. In pseudo-anonymisation, the data can still be indirectly linked to an individual. Therefore the data is still considered personal data. A example of pseudo-anonymisation is *key-coding* data. This is commonly used in medical research as pseudo-anonymisation is essential for safeguarding in research. Patient data is labelled with a code (‘*key-coded data*’) and that is used in medical research. It could be indirectly linked to the person if additional information is available.

A further way to safeguard personal data in research is the process of ‘data minimisation.’ Data minimisation is where data are released to the researcher that are *adequate* (sufficient to fulfil the research question, *relevant* to the research in question, and *limited* to what the researcher needs. Limiting the data for the researcher involves ensuring that the number of participants are minimised, the amount of data per participant are minimised and the degree of sensitivity is minimised.

Data governance is the process of standards to oversee the transfer, access, storage, use and disposal of data (including collaboration, sharing and preservation) of sensitive research data [508, 509]. There are four pillars of data governance when creating a solid data governance framework [509]. These are: data quality, data stewardship, data protection and compliance, and data management [509]. To achieve this, data governance of routinely collected health data is increasingly divided between different key roles. Therefore, there are a number of terms commonly used which are outlined in Table 3.1.

Term	Definition
Data Owner	This is the organisation/individual who are responsible for maintaining the data and ensuring outcomes are in line with the Data Governance Framework. They do not ‘own’ the data but are accountable for: classification, protection, use, and quality of the datasets. Data governance ‘outcome-focused’.
Data Steward	Assists the Data Owner. A person/group with expert subject knowledge on the dataset. They must ensure its quality, validity and appropriate use in line with agreed data governance. Their focus is on ‘tasks’ to manage the data and allow the data governance outcomes to be achieved. Data governance ‘Task-focused’.
Data Custodian	Responsible for safe custody, storage, and security controls for datasets to comply with the Data Governance Framework by the Data Owner.

Table 3.1: Definitions of common governance terms use in routinely collected data

The legal framework which governs personal data in the UK is The General Data Protection Regulations (GDPR) and the Data Protection Act (DPA) 2018 [510, 511]. These apply to personal data in general and also if used in research where it becomes sensitive research data. To ensure the law is followed, research on personal data must be undertaken within a secure and safe environment. These are often referred to as a ‘safe haven’ or other similar terminologies, the over-arching term is Trusted Research Environment (TRE) [498]. A TRE provides a legitimate basis for accessing data along with technologies to comply with the Five-Safes Framework[512]. The Five-Safes Framework are outlined in Table 3.2.

The Five-Safes	Description
Safe data	Data is treated to protect confidentiality
Safe projects	Research projects approved by data owners for public good
Safe people	Researchers are trained and authorised
Safe settings	A SecureLab environment for only those authorised
Safe outputs	Screened and approved non-disclosing outputs

Table 3.2: The Five-Safes Framework for research data - text extracted from [512]

In Chapters 4, 5, and 6 I discuss the specific processes involved to securely access and use the personal data in my research in Aotearoa New Zealand (ANZ) and Scotland and discuss the Five-safes framework [512] in relation to this. The data access environment differs across the four nations of the UK and internationally. Within England, Wales, and Northern Ireland, the TREs allow data to be re-searched in a database format whereby researchers use structured query language (SQL) to interrogate the database. This was similarly the process in ANZ, as will be described in more detail in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#). However, in Scotland there is a different process overseen by Public Health Scotland and the researchers are given version copies of national data files within the Safe Haven. This will be further discussed in the context of my research in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#).

Across all four nations of the UK, the legal basis for the governance model uses GDPR, Data Protection Act (DPA) 2018 and National Health Service Act 2006 Section 251 (S251) regulations rather than a requirement to apply through the Integrated Research Application System (IRAS) [510, 513, 511, 514]. A recent publication examining all 15 UK TREs highlighted variation in the rules and procedures on data access requests (DAR) which are unique to each TRE [498]. This review evaluated the need for standardisation in four key domains to ensure research can progress safely without unnecessary delays. These four domains aim to break down existing barriers and prevent segregation and are outlined in Table 3.3 adapted from [498].

Four key domains	Description
Data access environment	Harmonising the infrastructure and protocols for establishment of TRE computing systems
Data access requirements	Streamlining and aligning data access criteria
File out review procedures	Enabling uniform processes to review data extractions
Researcher accreditation/training	Promoting consistency required by TREs

Table 3.3: The four key domains to improve standardisation in the UK TREs - text extracted from Torabi et al. 2023 [498]

Additional aspects to the overall data governance, access and permissions include key terms such as ‘data stewards’ and ‘data custodians’ along with an essential aspect of researchers using any personal data which is ensuring Patient, Carer and Public Involvement and Engagement (PPIE).

3.7 Reporting routinely collected data research

In 2015 it was acknowledged that a specific structure for reporting research using routinely collected healthcare data was essential. As a result, the Reporting of studies Conducted using Observational Routinely collected health Data (RECORD) statement was published [433]. This was produced as a recognition that existing general reporting guidelines for epidemiology studies such as Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [515] lacked key areas that needed to be addressed when publishing research using routinely collected health data [433]. One important aspect that is addressed is what is classed traditionally as ‘*study* population’ within epidemiological research has additional layers to consider in routinely collected health data. This includes the need to report the *source* population, *database* population, and *study* population as demonstrated in Table 3.4. The full RECORD checklist [433] is in Appendix 2. There are other key aspects to consider in relation to the differences between STROBE and RECORD guidance in publishing research methods and results using routinely collected health data.

Population hierarchy in studies using routinely collected data sources	
<i>Source</i> population	Population from which the database is derived and from which the researchers want to make inferences
<i>Database</i> population	Population derived from the source population and comprised people included in the data source
<i>Study</i> population	Population identified from within the database population by researchers using codes and algorithms

Table 3.4: Population hierarchy in studies using routinely collected data - information extracted from Benchimol et al. (2015) [433]

The population hierarchy in routinely collected data from Benchimol et al. is also clearly illustrated in Figure 3.1 (Population hierarchy in studies using routinely

collected data sources) [433]. The importance of this figure is the change from the larger *source* population, to *database* population, to actual *study* population that the researcher uses for analysis [433].

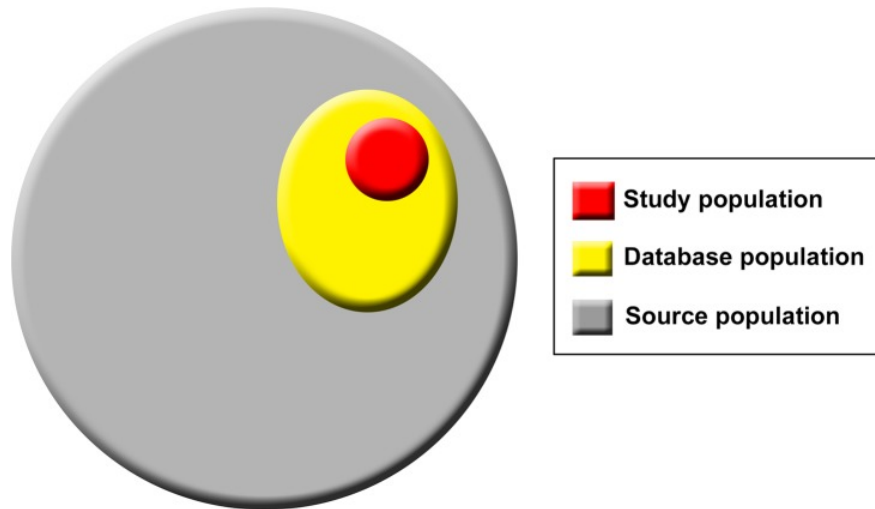


Figure 3.1: Population hierarchy in studies using routinely collected data sources [433]

Source: Reproduced from PLOS Medicine, Vol 12(10): e1001885. Authors: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al., Title: The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement, 2015. Used with permission from PLOS Medicine under open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are properly credited.

Since the RECORD guidelines were published [433], there have been further extensions of these for publishing routinely collected health data research focusing on pharmacology (RECORD-PE) and trials (CONSORT-ROUTINE) [516, 517].

In addition, it is important that research including routinely collected health data is clearly identified as such, particularly to allow assimilation of knowledge and identification of appropriate research for systematic reviews. Currently, there are no accepted Medical Subject Headings (MeSH) for identifying studies using

this type of health data. To identify this type of research then carefully selecting search terms that capture the range of terms used for routinely collected data/electronic health data research is required. This is to prevent research being missed. The RECORD guidelines [433] specify details that must be included in the title of a publication which can help with this, but often authors still use a range of different terms. Having an agreed MeSH heading for routinely collected health data research would be valuable.

3.8 Routinely collected data: dementia

In [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#), [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#) I discuss the different datasets used within routinely collected data for ascertaining dementia. There has been much work undertaken assessing validity and triangulation of datasets for ascertaining dementia. It is recognised that the most common datasets available within routinely collected data will often miss people who may have dementia. Such people who do not show up in routinely collected datasets include people who do not need (use as often) secondary care services (e.g. hospitals) and therefore do not feature in the corresponding datasets as often or at all, are not able to have a dementia drug prescribed due to contraindications, or when they die have other conditions listed on their death certificate rather than dementia. It is now well recognised that including multiple datasets [518] increases the validity of ascertaining dementia and also allows triangulation of results as further discussed in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#).

Another systematic review (2018) investigating the identification of dementia cases with routinely collected health data identified 25 studies estimating Positive Predictor Value (PPV) and eight estimating sensitivity from the eligible 27

studies [152]. Whilst the authors recognised that the study settings and methods varied widely, they found that for all-cause dementia, PPVs ranged from 33%-100%, for 16/27 studies the PPVs were >75% [152]. Of the 8/27 studies investigating sensitivities, the systematic review found these ranged from 21% to 86%. The study found that the PPVs for Alzheimer's disease (range 57%-100%) were generally higher than those for vascular dementia (range 19%-91%). This systematic review shows that if someone had a diagnosis of dementia in routinely collected health data, this is likely to be accurate. There have been other systematic reviews investigating the validity of ascertaining other neurological diseases from routinely collected data with similar results [519]. Other countries have also assessed the validity of using routinely collected health data for dementia ascertainment [520].

In the UK, primary care is typically the first point of contact for any individual who has concerns regarding their memory and may have or be at risk of having dementia. Using primary care datasets have been found to be very valuable in ascertaining dementia, although some information may not be coded in a diagnostic format and valuable information may be found in the written text of consultations. This is an ongoing area of 'big data' in healthcare that is being addressed. e.g., the move from READ codes in primary care to SNOMED CT codes will enhance the ability to link primary care data and therefore increase the ascertainment of dementia using routinely collected data. In this section I discuss the three datasets that I use within this thesis, with further detail discussed in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) and [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#). One important aspect to highlight in this section is the need for researchers consistently using the same set of codes for dementia, and the recognition that with time the codes may change or clinical recording of codes improve. This may be influenced by national drives for dementia diagnosis, or as with other diseases, the possibilities for an effective treatment. This was seen in coding of stroke diagnoses as the recognition and management became more es-

sential as stroke thrombolysis treatment emerged and there was ‘something that could be done’ and a coordinated action plan with a public awareness campaign.

3.8.1 Death certificates

Death certificates (The Medical Certificate of the Cause of Death, MCCD) in the UK are composed of two parts which are completed by a registered medical practitioner [521, 522]. Part 1 should refer to the immediate cause of death i.e., the disease or condition that led directly to the death or any intermediate causes of that disease or condition and the rest of part 1 should highlight the chain of events leading to part 1 (i.e., "*(a) the disease or injury which initiated the chain of morbid events leading directly to death or (b) the circumstances of the accident or violence which produced the fatal injury*") [521].

Part 2 is used to highlight other significant diseases, conditions or accidents which may have contributed to death [521]. It should not include all the conditions that were present at death [521]. Someone could have dementia as an ‘other condition’ but not have it registered in part 2 of the death certificate. Equally, if the clinicians felt that dementia contributed to death due to a reason e.g., increasing frailty, then it may feature in part 2. Another important aspect is that up to ten codes can be recorded on Scottish death certificates. Therefore, even if the certifying clinician entered more than ten conditions in total across all parts of the death certificate, these would not be included.

Whilst it is a valuable tool within routinely collected data, death certificates clearly have limitations. The variation in reliability and accuracy of death certificates as a source for ascertaining dementia can vary substantially from 50% to approximately 70-80% in Scotland. However, there have been a number of studies looking at the accuracy of dementia entries in many countries. Another paper looked at the reliability of death certificates for surveillance for dementia at a population level within the Medical Research Council Cognitive Function and Ageing Studies (CFAS) and CFAS II studies) [523]. Importantly, this paper also

looked at the time trend of re-coding dementia on death certificates. This will vary and indeed with COVID-19 there were further changes to death certification that might affect this [524]. My research focuses on routinely collected data up until 2018 in Scotland so is not affected by those recent changes.

The gold standard would be to triangulate death certificate recording of dementia in those who attended a memory/dementia clinic in all routinely collected data because by default their diagnosis should be valid. However, this is often not possible. Death certificates are an accepted source in routinely collected data for ascertaining dementia, but with the need to increase the number of sources for dementia ascertainment to increase accuracy [518].

Dementia may feature at any point in the death certificate, but is more likely to be found in part 2. Therefore, it is important as a researcher to interrogate all parts of the death certificate data to ensure a diagnosis of dementia is not missed and not rely on the underlying cause of death.

3.8.2 Hospital discharges

Another study looking at the accuracy of general hospital dementia diagnoses found that diagnostic sensitivity increased between 2008 and 2016 [525]. This is important given this may reflect many changes at a population level, including the drive from the G8 Dementia Summit and increased government targets for dementia and subsequent population awareness driving up the desire for many to seek a diagnosis and for clinicians to be more likely to make a diagnosis. Hospital admission datasets vary across the UK. In Scotland there are a number of Scottish Morbidity Record (SMR) datasets covering Outpatients, General/Acute Inpatients/Day Cases, Maternity and Mental Health. They all use ICD codes for coding diagnoses. In New Zealand there is a very similar set up which will be discussed in more detail in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#).

As discussed earlier in Section 3.4 ([Healthcare data: types and coding](#)), there are subtle variations in the ICD versions around the world and over time, codes for new diseases or better definitions of conditions, including sub-types of dementia can change. In [Chapter 1: Introduction](#) I discussed the different methods for diagnosing dementia, DSM criteria and ICD are key methods. Interestingly, a recent study from Sweden investigating routinely collected data found that ICD version 11 had the highest yield of dementia prevalence in their study followed by DSM-5 and DSM-IV in relation to dementia [526]. This highlights that the choice of diagnostic criteria is important. As the UK, and international community, move to recording diagnoses using the SNOMED CT criteria it will be interesting to examine how this influences dementia incidence and prevalence within hospital datasets.

There are a clear set of ICD codes (ICD 9, 10 and 11) that relate to dementia diagnoses [145]. This means they can be used when interrogating the hospital admissions databases to ascertain who had dementia recorded on their discharge summary. Frequently, the ICD codes used for research can have a temporal lag to newer versions of ICD codes if prior ICD codes have to be ‘mapped’ to newer ICD codes for a disease to ensure continuity. It is very important that if dementia is *not* recorded (in either the main condition, or other conditions option) that it does *not* mean the person does not have dementia. Particularly in acute admissions where there may be another acute reason for their deterioration (e.g., pneumonia) this would feature rather than dementia and dementia may not be recorded at any point on the discharge summary.

Research has been undertaken to investigate the reliability of hospital datasets for ascertaining dementia [527, 152, 528]. In England, a study investigating dementia diagnosis in primary care and hospital care found 26,269 people (55.4%) in primary care had a dementia diagnosis out of the total 47,386 with dementia in their study. This compared with 31,034 people (65.5%) with a diagnosis of dementia in hospital records and 12,232 people (25.8%) with a diagnosis of

dementia in mortality records [529]. Additionally, there was corroborating evidence for those with a diagnosis of dementia in primary care for 23,225 people (88%), 23,658 people (76%) in hospital records and 10,191 (83%) in mortality records [529]. This highlights that if a diagnosis of dementia is recorded then the likelihood of that diagnosis being correct is quite high.

Furthermore, the study showed that whilst there was overlap between the three data-sources for a diagnosis of dementia, only 8% (3,861 people) were found in all three datasets (primary care, hospital admissions and mortality registry) [529]. The dataset that added the most individuals with a diagnosis of dementia only found in that single dataset was the hospital admissions dataset (15,121 people, 32%) and the lowest was the mortality registry with 3,056 people (6%) with a diagnosis of dementia in only that dataset [529]. This confirms, what systematic reviews have shown, that using increased datasets within routinely collected health data increases the yield for dementia ascertainment.

More recently, in 2023, a study [530] investigating dementia ascertainment within the Lothian Birth Cohort 1936 in Scotland linked to electronic health records investigated self-reporting of a dementia diagnosis compared with documentation of cognitive dysfunction in electronic health records, and an expert consensus as part of a diagnostic review board of clinicians. The self-reported dementia diagnoses were positive in 17.8% of ascertained dementia diagnoses in the cohort [530]. The anticipated all-cause dementia prevalence from the expert clinicians (diagnostic review board) was similar to individuals identified as having cognitive dysfunction in the electronic health records or home visits (13.6%, 118/865 people from the diagnostic review board compared with 18.8% ascertained as having cognitive dysfunction in electronic health record reviews or home visits) [530].

This supports the view that in Scotland electronic health records can be a useful tool for finding those who potentially have a diagnosis of dementia, and whom expert clinicians would diagnose as having dementia from detailed assessment of

the records. Another important result from this study showed that of the 64 participants who died, 47 people (73.4%) had dementia recorded on their death certificates [530]. It is known that there is variation in a diagnosis of dementia being recorded on a death certificate. This study [530] highlights that it can provide a valuable resource for ascertaining dementia if additional datasets are also used to increase ascertainment.

In summary, it has been shown that there is a high positive predictive value (PPV) for dementia in electronic health records [152]. This highlights that if someone has a dementia diagnosis recorded in these datasets, then they are likely to have the diagnosis [152]. However, it is recognised that a number of people are ‘missed’ who have the diagnosis but don’t have it recorded at discharge. The sub-type of dementia is most accurately recorded in psychiatry settings than other general settings. Whilst Alzheimer dementia (Alzheimer disease dementia) is the most common sub-type recorded in routinely collected health data (electronic health records), it can be difficult to truly quantify sub-types due to the accuracy of sub-type dementia diagnosis and that multiple sub-types can be given (which will be discussed in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) with my results in Aotearoa New Zealand).

3.8.3 Prescription of anti-dementia drugs

There are a number of drugs that are used specifically as dementia drugs and this means they can be reliably used to ascertain dementia within prescribing datasets. These include acetylcholinesterase inhibitors (Donepezil, Rivastigmine, and Galantamine) and Memantine Hydrochloride [531]. The British National Formulary lists their (Donepezil, Galantamine, Memantine) licence use for mild-moderate dementia in Alzheimer disease dementia, whilst for non-AD they are prescribed off licence [531]. Rivastigmine is licenced for Dementia with Lewy Body [531]. Therefore, people with a range of sub-types of dementia may be on these drugs. They are often referred to by their trade names but for ascertaining within datasets the chemical name listed above is used. The most common de-

mentia drug prescribed is Donepezil. Memantine is licensed for moderate-severe AD [531]. It is important to recognise that not everyone with dementia will be prescribed a dementia drug if they a) don't wish to be or b) have contraindications to its prescription meaning it would be inappropriate or c) experience side effects from the drug.

In summary, it is possible (and has been validated) to use common datasets available traditionally within routinely collected healthcare data to ascertain dementia. I use this approach in Chapters 4, 5 and 6 of this thesis, for research in Scotland and Aotearoa New Zealand.

3.9 Routinely collected data: geography

Much big data are geo-coded – i.e., they incorporate location or some other geographical information. As individuals we are often geo-coded, without being aware of it, for instance through applications (apps) on our smart phones or by using the internet that geo-codes our IP address. Within routinely collected data, geographical data have always been collected in some form. For instance, the location (city/building/address/name) of the healthcare facility that a person attends and the address of the patient are common geographical data collected. This is important to provide healthcare. For example, when healthcare staff are checking that an individual is the 'correct' patient for an appointment or procedure then the three common person identifiable pieces of data that are checked are name, date of birth and address.

Geographical data in routinely collected health data also allow access to primary care (such as General Practice, GP) services in the UK as an address is needed to access GP services for permanent residences. Geographical data also allows access to healthcare services nearest them within a geographically defined NHS health board area, and through that funding for NHS health boards are based on the numbers of individuals within a geographical area that the health board

must provide services to. In Scotland the patient's address is transcribed from their registered GP practice and linked with their CHI-number.

The difficulty of using geographical data in routinely collected health data for research is that it can be a further personal identifiable piece of data that can identify a person – directly or indirectly. There are many different levels of geographical data from small areas to large geographical areas. Geographical data can be available at individual level (i.e., the person's specific address), small areas (e.g., data zone or postcode), or larger areas (e.g., council area, county, or health board. The smaller the area (smallest being at an individual level) means the data is increasingly identifying. Using a small geographical area does not mean that de-identified people within routinely collected healthcare data will automatically be identifiable, if the disease being investigated is highly prevalent within that area. The more the researcher narrows the research question (e.g., by age or sex or other specifics), then the potential to identify someone increases.

This is obviously greatest in a rare disease where the prevalence is very low, so that even in a large geographical area it may still be possible to identify the individual indirectly. Therefore, including geographical data in healthcare research using routinely collected healthcare data can make those governing the data access cautious. Ways to mitigate this whilst protecting individuals' personal data are to use the smallest geographical scale that allows the least risk of being identifiable. Figure 3.2 reproduced with permission from [532] below highlights what is referred to as the privacy-access continuum framework. This summarises the level of access to data within routinely collected healthcare data, the geographical scale, and the potential for identification. Subsequently, in the book chapter [4] I co-wrote on routinely collected data for healthcare research during my PhD, we also make reference to this important framework.

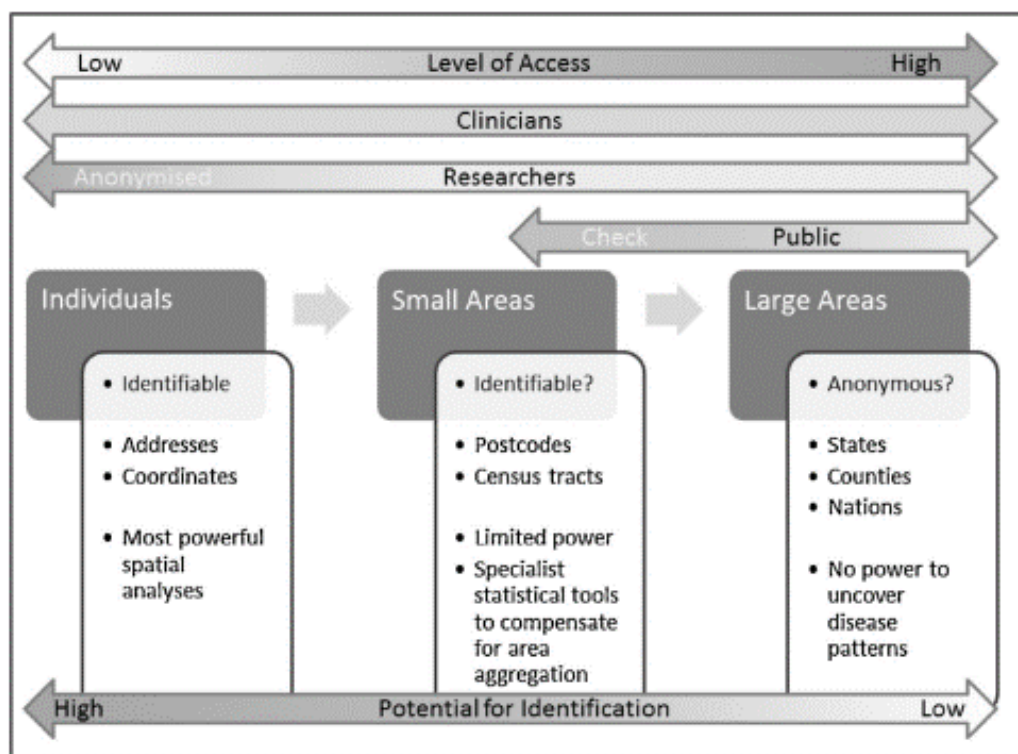


Figure 3.2: The privacy-access continuum framework [532]

Source: Health Policy Journal. Authors: Daniel John Exeter, Sarah Rodgers, and Clive Eric Sabel. Title: "Whose data is it anyway?" The implications of putting small area-level health and social data online. Copyright (2014). Used with permission from Elsevier and Copyright Clearance Center.

Since the COVID-19 pandemic, the WHO has a drive to share data between and within countries, this includes geo-located healthcare information. There is recognition that even within High Income Countries there was a paucity in information understanding where healthcare facilities were, and what different areas could provide during the pandemic. This led to the development of the WHO Geolocated Health Facilities Data (GHFD) initiative – the geographical element to delivering healthcare [507]. The WHO now have a wealth of resources and meta-data packages for countries to assist with unifying the collection and analysis of routinely collected health data. When considering geo-coded information at an individual level there are still challenges within routinely collected health data

of ensuring it does not identify an individual. However, there is clearly a need to understand the role that location plays in our health, well-being and diseases as outlined in [Chapter 1: Introduction](#) of this thesis.

The UK is now approaching this in an interesting and exciting way with the equivalent of a Unique Identifier for geographical data. At the end of 2023, the UK government published a key report from the Geo-spatial commission that was previously established in 2018 [\[533, 534\]](#). This report was entitled: ‘UK Geo-spatial Strategy 2030, Unlocking the power of location’ [\[533, 534\]](#). Understanding location in a person’s life journey is important. Having geographical data that are meaningful and usable within routinely collected health data, without identifying someone, is powerful for furthering understanding of health and diseases inequalities and variations. The proposal is to use a Unique Property Reference Number (UPRN) which the UK Ordnance Survey produces [\[534\]](#).

The UPRN is a unique identifier for every addressable location in Great Britain [\[379, 534\]](#). Separately there is another unique geo-spatial identifier called a USRN (Unique Street Reference Number) which is for every street in Great Britain [\[534\]](#). Both of these, along with their associated grid references are available under an open data licence from the Ordnance Survey Data Hub. The UPRNs are able to aid in integrating and sharing geo-spatial data across different systems particularly between local authorities, emergencies services, and central government [\[379, 534\]](#).

In terms of their use in healthcare, there has been recent research undertaken investigating UPRN in routinely collected health data in NHS England [\[535\]](#). Research has recently also begun investigating using this in NHS Scotland [\[536\]](#). The CHI-number in Scotland is linked to the patient’s address registered with their GP [\[482\]](#). The research in Scotland is investigating the accuracy of the UPRN seeded in the address linked with a person’s CHI number within routinely collected healthcare data in South-East Scotland’s DataLoch [\[536\]](#). If successful,

this would remove the need for addresses and reduce risk of identifiable geographical data. It could prove a powerful and important way of investigating location, health and disease within routinely collected healthcare data in the future.

3.10 Health data research in relation to ageing and dementia

3.10.1 UK health data landscape

Within the UK, across the four nations, there are broadly similar datasets available – primary care, secondary care, prescribing datasets within healthcare, and mortality datasets via national administrative centres – but specific details vary between nations.

As discussed earlier in this chapter routinely collected health data vary across the UK in both how it is linked and accessed with no uniform system across the four nations. An independent UK wide review was established in 2023, ‘The Sudlow Review’ led by Professor Cathy Sudlow looking at ‘Unifying Health Data in the UK’ across the four nations [537] and is expected at the end of 2024. The aim of the review is to *"evaluate how data can be better managed to improve public health while maintaining privacy and trust"*. This will be a valuable and much needed independent review to progress health data research in the UK. I contributed to this review as part of the Ageing Data Research Collaborative research group.

This is on the background of other drives for digitalising health records, creating NHS bench-marking networks and digital innovation hubs to improve the ability to use health data meaningfully, and in a timely way, in data research. The Health Data Research UK (HDR UK) is the national institute for health data science [538]. Established in 2018, the aim is to *"unite the UK's health data to enable discoveries that improve people's lives"* [538]. It has three goals to achieve this: 1) accelerate trustworthy data use, 2) empower researchers, 3) promote partnerships.

One of their initial programmes was the 2018 Digital Innovation Hub Programme [538]. This aimed to establish a national infrastructure for health data science. The programme established three components which interlinked that they felt were key to a successful national infrastructure for health data science. These are:

- The Health Data Research Hub Network (the Hubs)
- The UK Health Data Research Alliance (the Alliance)
- The Health Data Research Innovation Gateway (the Gateway)

The hubs are collaborations that span NHS, industry, academia and public and patient stakeholders, and work to support different disease projects [539].

There is real scope for the role of digital data in the healthcare of older people as highlighted in a recent review article [6] which I co-authored with members of the Ageing Data Research Collaborative. However, there are still challenges as this article outlines [6]. One challenge is that approximately 45% still use paper records within social care data [6]. This, along with the other challenges, are important to be mindful of as a researcher using health data. It highlights that the outputs from health data research using routinely collected data may be a smaller view of the larger picture not yet digitalised.

The variation in health big data across the UK does need addressed for future development in research across the four nations. Data And Research Environments (DARE) UK is also addressing this with the *"aims to design and deliver a coordinated and trustworthy national data research infrastructure to support cross-domain research for public good"* [540]. DARE UK is a collaboration between three health research groups in the UK: UK Research and Innovation, Health Data Research (HDR) UK, and Administrative Data Research (ADR) UK [540].

3.10.2 Current Scottish data strategy

As outlined in Section 3.5.1 (Unique identifiers), Scotland has the CHI-number in healthcare which is a unique identifier. There is currently a move in Scotland underway to a new CHI system [541]. The new CHI system became live at the end of 2023 and is embedded into the NHS systems. The potential benefits of this are threefold:

- *"potential to move away from the current numbering format so that personal information is no longer contained within CHI number which would provide better patient confidentiality"*
- *"better infrastructure, the new system is capable of holding multiple addresses so patient data can be updated from multiple sources, improving the accuracy of patient data"*
- *"24/7 functionality" [541]*

Additionally, with the drive for merging health and social care in Scotland, there is work underway to explore the benefits and risks of using this new CHI system in Local Government with the potential to *"act as a verification tool for accessing the Digital Front Door and the Integrated Social Care and Health Record"* [541]. This could be powerful in allowing further data linkage than the current CHI-system allows, as a purely health driven UPI.

Other changes since my published review in Section 3.2 (Peer-reviewed published manuscript), include changes in data access, governance, and data strategies in Scotland. All the key centres for health data research in Scotland mentioned in the article have now changed including Information Services Division (ISD), a division of National Services Scotland and linked with NHS Scotland, which is now part of Public Health Scotland (PHS) [542]. PHS now has responsibilities as a ‘data controller’ in Scotland [542]. This also reflects what is happening in other parts of the UK and the importance for collaboration and less working in silos as well as changes in centres’ names as funding, and remit of the centres, changed with time.

In place of Scottish Informatics Programme (SHIP), Farr Institute of Health Informatics Research, Administrative Data Research Network (ADRN), Scottish Informatics and Linkage Collaboration (SILC) we have the Scottish Centre for Administrative Data Research (SCADR), Health Data Research UK (HDRUK) and Scottish Government Linkage Projects. All of these are supported by the electronic Data Research and Innovation Service (eDRIS) team, part of Public Health Scotland [542]. The Research Data Scotland (RDS) system began in 2021 to review how to streamline the service for accessing routinely collected data in Scotland for researchers [543]. eDRIS is a part of RDS [543, 542].

Data requests normally require approval through the PBPP Approval Pathway in Scotland for routinely collected data. There are two Public Benefit and Privacy Panels: health and social care (HSC-PBPP) and Statistics (S-PBPP) [544, 545].

In 2021, Scotland produced a key document called ‘Care in the Digital Age’ [546]. This is a dedicated Data Strategy for Health and Social Care, building on the first Digital Health and Care Strategy 2018, and is nested within the overall ‘Digital Strategy for Scotland’ [547, 546]. Figure 3.3 below outlines the eight principles of a digital nation from this government report. As with many other aspects of health data innovation this recent strategy was also driven by the COVID-19 pandemic demonstrating the important role of health data. The vision is *“to improve the care and well-being of people in Scotland by making best use of digital technologies in the design and delivery of service”* [546], ultimately ensuring a ‘right care, right place, right time’ approach.

One priority is to make better use of the data that are already held, but also data that are not currently routinely collected/held [546]. Priority three within the document discusses the ‘Digital Foundations’ and the development of the ‘National Digital Platform’ (NDP) highlighted in the first strategy document [546]. Alongside this platform is the need for a common shared single language within the clinical and care IT systems. SNOMED CT is a clinical coding system

mentioned earlier in this chapter for use in electronic health records and will be used across Scotland [475, 548]. SNOMED CT will theoretically allow exchanging information between systems [475, 548].

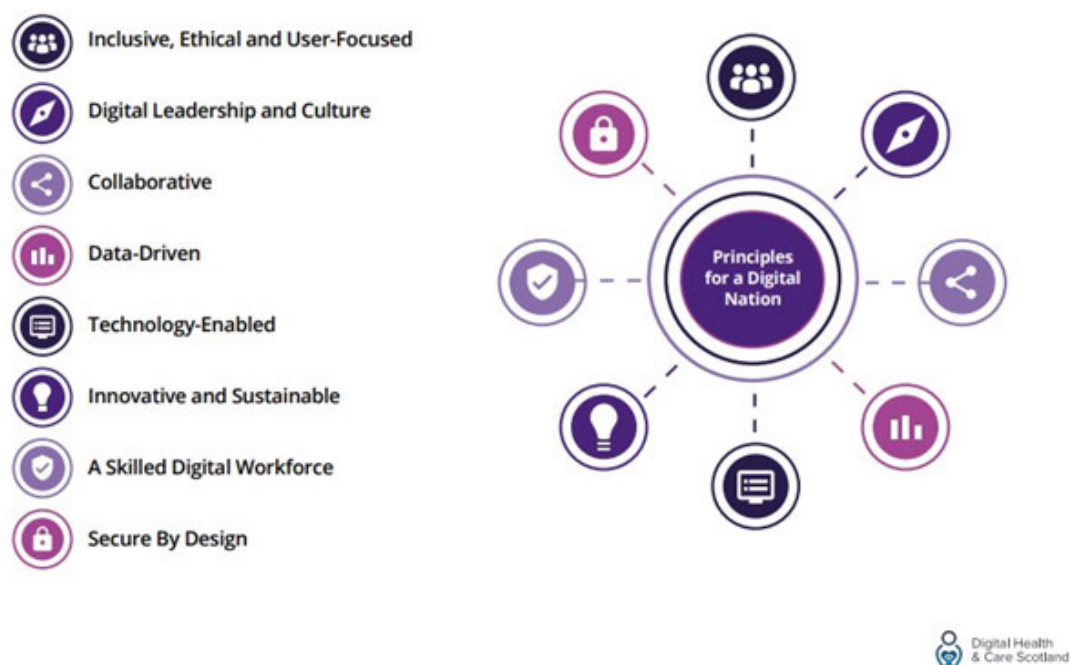


Figure 3.3: Scottish Government’s 8 Principles of a Digital Nation [546]

Source: Enabling, Connecting and Empowering: Care in the Digital Age report, Authors: Scottish Government and COSLA, Edinburgh, 2021. Used with permission under Open Government Licence Version 3 available at: <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

Health data research and creating an integrated health and social care system, including data, are constantly evolving with additional developments and updates since the ‘Care in the Digital Age’ document in 2021 [549, 550]. This includes work to bring together health and social care data, and professionals, for a holistic data-driven approach. There are numerous health data driven innovation groups across Scotland, with each region having its own Scottish Trusted Research Environment (TRE) for bringing University and NHS professionals together. In South

East Scotland, DataLoch [536] builds on a similar already established group in Dundee, the Healthcare Informatics Centre (HIC) [551] which combines health and social data and was established over fifteen years ago.

Alongside this there is ongoing work to collaboratively improve health data access through a project called the SARA project [552]. This is a collaboration of academic partners and TREs which is funded by UK Research and Innovation as part of the DARE UK programme [540]. The aim of the SARA project is to enhance the landscape of health data by improving the trustworthiness of data provenance, and also to minimise the risk of identifiable information within clinical free text records which hold sensitive data [552].

In 2021, the Scottish Government established the Research Data Scotland (RDS) service [543]. It has been a service that has largely been planning and reviewing the linked data options in Scotland and how to facilitate access for researchers in line with the Five-Safes Network. The first significant change came in April this year (2024), when Research Data Scotland (RDS) [553, 543], alongside eDRIS, co-developed and released a new service for researches called the *Researcher Access Service*. The aim of this is to make it faster and simpler to access secure public sector datasets for research. This includes a streamlined approvals process and at the initial launch, nine of the most frequently used Public Health Scotland's datasets are available to researchers using this streamlined approvals process. These include the datasets such as SMR 01 (1981 onwards), SMR 04 (1981 onwards), PIS (from 2009 onwards), and Deaths (1981 onwards) that I used in my research in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) [553]. As the service develops, RDS is working with other organisations to widen the number of datasets available. The new concept with the RDS is to be designed specifically around the needs of researchers and data uses. It is closely aligned with the principles of the Scottish Approach to Service Design and builds on the work that eDRIS has done and its existing structures and services. This could provide an important step in using routinely collected health datasets

in Scotland for researchers and streamlining, in a timely way, the processes for accessing the data for research.

3.10.3 Developing ageing health data research UK

Given the UK's ageing population and complexity of their conditions, there is a need for a holistic approach to research of the older person using routinely collected health and social care data. In 2021, the Geriatric medicine Getting It Right First Time (GIRFT) report in NHS England highlighted widely varying practices in how geriatric medicine and frailty data were used between health-care organisations [554]. People with dementia will often have multiple chronic conditions and frailty. This means there needs to be a holistic multi-disciplinary approach to research for people with dementia and this need does not change because it is routinely collected data. Therefore, for developing a future ageing health data research ready UK, it is even more important that the routinely collected data are not solely health data but health and social care data.

3.10.3.1 Digital passport and shared care record

In the peer-reviewed publication I co-wrote [6], we highlighted how a digital passport or shared care record are examples of including the patient in their electronic health journey. There are a number of examples where shared care electronic health records have been a positive feature. This includes 'My Health Record' in Australia, and also systems such as Sci-Diabetes in East Scotland, or MyGOSH (which is part of MyChart by Epic and is an app, and webpage, for Great Ormond Street Hospital for patients and clinicians to communicate information, blood test results, investigations, outpatient/inpatient letters) [555, 556, 557, 558]. These examples allow the patient to interact with their electronic health records. Some of the shared care records involve a single healthcare speciality (e.g., diabetes or renal services), and some span all hospital interactions.

With an ageing population that requires holistic care, and often input from different specialities, it is important to have harmony of services. For instance, a

condition such as dementia falls within different healthcare specialities (e.g., old age medicine, palliative care, psychiatry, primary care, general medicine, physiotherapy, occupational therapy etc.) alongside other care providers (including social care) who also need to contribute care to an individual patient. An integrated approach is ideal. Whilst patient access to the data is also important, many patients may need support to access their records. This reinforces the need to holistically consider the individual interacting with health and social care data on their ‘patient journey’ to allow a successful digital shared care record. For this to happen there needs to be people with understanding of the complexity of ageing and dementia involved in the future digital transformation of routinely collected health and social care data. This would enhance future health and social care data research.

3.10.3.2 Ageing Data Research Collaborative

In 2018, the Ageing Data Research Collaborative (ADRC) was established [559]. The principal aim is to *“to promote big data research among the geriatric medicine community, and work to increase the dialogue between data science and clinicians particularly in the field of ageing research”* [559]. It was developed following recognition there was a specific need for those with an expert in ageing research to help drive forward health data research for older people in the UK. Since then, several publications, workshops at the national British Geriatrics Society and establishing a Specific Interest Group (SIG) in European Geriatric Medicine Society (EUGMS) have been achieved [560]. It is recognised that a multi-disciplinary approach is required to undertake ageing data research and develop a future ageing health data research ready UK.

3.10.3.3 Minimum dataset

The Professional Record Standards Body (PRSB) has created guidance on minimum standards for care records [561]. Whilst they work across the four nations, only NHS England is a member [561]. The PRSB also includes a person-focused section whereby an individual can record what matters to them for communica-

tion with those providing support and care [561]. What constitutes a ‘minimum dataset’ for ageing data research is complex. Research has investigated what would constitute a minimum dataset (MDS) within care homes as part of the Developing research resources And Minimum data set for Care Homes’ Adoption and use (DACHA) study [562]. Further work has investigated what could be classed as ‘realistic’ for an MDS in this setting [563]. This will be piloted in a feasibility study in areas of England [564]. It is important to be able to capture where people live, including those in care homes. There can be movement at the end of life when frailty increases with nursing home requirements. To be able to investigate geographical locations from a longitudinal perspective in datasets, knowing where people live across their life time is important and also if the movement at the end of life is due to the location of care home. However, as highlighted both before COVID-19 and subsequently, there are great difficulties in identifying those in care homes within health datasets [565, 497]. It is known if dementia ascertainment is taken solely from mortality datasets then the location can reflect a care home. This can introduce bias in interpreting location and dementia diagnoses using routinely collected health data as those with complex care needs including severe dementia are more likely to require care that cannot be managed at an individual’s home location.

I was involved with other UK geriatric medicine researchers to create a Minimum Data Set (MDS), or core dataset, for geriatric medicine research [11]. However, this is not specific to routinely collected health data research. Further discussion and research are needed to consider whether a MDS for ageing data research using routinely collected health data are appropriate for developing a future ageing health data research ready UK. Given the growing interest in potential modifiable risk factors for dementia and the role the environment (and location) may have across our life course in influencing health and disease, it is important that geographical data is also captured in minimum datasets for routinely collected data in ageing research.

3.11 Chapter conclusions

In summary, big data in healthcare in the form of routinely collected health data can be repurposed for a secondary, and valuable, use for healthcare research. There is a large amount of routinely collected health data that can be used for dementia research. Using routinely collected health data has challenges and limitations that are different from traditional epidemiological approaches to investigating dementia. COVID-19 has created a drive to enhance data linkage, sharing data and increased awareness of the benefits of epidemiological approaches and observational data with the public, researchers and policy makers for understanding disease variation.

My peer reviewed published study showed that there was potential to create a linked health and social care resource for healthcare research in Scotland due to the unique identifier (the CHI-number) in Scotland although there were challenges in linking health and social care. Since my paper was published in 2017 [2], there has been a growth in the recognition of the role routinely collected health data can have in healthcare research. COVID-19 has highlighted the potential for collaborative work in real time to inform healthcare delivery and research into a disease. There still remains many challenges in the variations of TREs across the UK alone, and the way data are accessed and linked. There are additional complexities in linking data between countries (including the four nations of the UK) which have different systems and unique identifiers. However, there are European, UK and WHO initiatives to investigate ways to navigate the barriers.

For dementia, a longitudinal approach (life-course) approach as discussed in Chapter 1 is important for understanding this complex disease with a holistic approach. That can be challenging within routinely collected health data. However, integrating other datasets (e.g. cohorts, surveys etc) with routinely collected health data provides a rich resource for researchers. This is possible in a quicker and more efficient way if individuals have one unique identifier for all healthcare

interactions, such as the CHI-number in Scotland. In some countries there are unique identifiers that span health and administration.

As the individual four nations develop their approaches to integrate systems, including using a coding system such as SNOMED CT that spans both primary and secondary care, there is hope that increased integration will prevent silos and encourage data linkage which will benefit healthcare research and the people whose data it is whilst also respecting the Five-Safes Framework of using routinely collected data for this purpose. There is still variation in the way researchers access data across the UK for research using routinely collected health data with variations in the processes, applications and governance structures. There are positive steps (including the four domains to address) with discussion to consider a more uniform approach. The Sudlow Review: ‘Unifying Health Data in the UK’ [537] will provide an interesting report on the future of health data use in the UK as routinely collected health data, even more so following COVID-19, has become an increasingly valuable resource for epidemiological research into dementia.

I now explore using routinely collected health data in my subsequent [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#) to investigate my research question: *"Does dementia vary between rural and urban areas, and how can we use routinely collected health data to answer this question?"*

Chapter 4:

Investigating dementia in Aotearoa New Zealand (ANZ) using routinely collected data

4.1 Chapter introduction

With a 2023 estimated population of 5.22 million, Aotearoa New Zealand (ANZ) is a country of similar population size to Scotland's 5.5 million in 2023 [566]. There are also similarities to Scotland with five main cities (Auckland, Wellington, Christchurch, Dunedin and Hamilton), and smaller townships surrounded by a rural environment. I refer to New Zealand (NZ) as Aoteaora New Zealand (ANZ) in this Chapter. Since my published manuscript on my research in ANZ, there has been a growing movement in academia to correctly refer to NZ as ANZ, recognising its culture and Indigenous Population in the country name.

Whilst there are many differences between the two nations (Scotland and ANZ), one important similarity relevant to my research, is the similar shared structure of a national health service and ability to link routinely collected healthcare data. There are of course some important differences in how the national health services are delivered and how data are linked in the two nations.

To develop country-specific health policies that meet the healthcare needs for those with dementia, as well as ensure healthcare services are designed for the population they serve, there must be an understanding of the country-specific burden of disease and prevalence of dementia. Prior to visiting ANZ to undertake my research (and initial scoping exercise in 2016) I had not been aware that

the country specific estimates for ANZ were not based on their own population data. Despite the clear need to have nationally representative data that accounts for the diverse population groups within a country, the dementia prevalence estimates for ANZ were based on other population’s data. Indeed, the ANZ estimates of dementia prevalence are based on EURODEM prevalence data [77], some additional Australian dementia prevalence data, which were extrapolated to reflect ANZ’s population demographics.

The estimates published in 2021, estimated the number of people in ANZ with dementia to be approximately 70,000 people and was expected to reach just under 170,000 by 2050, compared with 90,000 in Scotland in 2024 and 55 million worldwide in 2050 [567, 92, 568]. It is known that these estimates may be less than the actual prevalence rates in both countries. As highlighted earlier in [Chapter 1: Introduction](#), diagnostic rates for dementia can be lower (i.e., those who are diagnosed with dementia do not represent the ‘true’ picture of all who may have dementia), including in High Income Countries (HICs) where it is estimated that only approximately 40-50% of people with dementia receive a diagnosis [92]. Therefore, the prevalence of dementia in ANZ may be higher, particularly as the 2021 prevalence estimates relied on international estimates, to reflect its unique multi-ethnic population.

4.2 Research aims and objectives

The aims of this study were to use routinely collected data in the Statistics New Zealand’s Integrated Data Infrastructure (IDI) to:

1. Investigate ANZ’s late-onset dementia prevalence from 2012 to 2015 (inclusive);
2. Compare this with published ANZ estimates; and
3. Compare variations between the North and South Islands and ethnicity

This chapter begins with my published peer-reviewed research from ANZ investigating dementia prevalence and geographical variation between North and South Island [3]. To undertake this research, I used a national routinely collected data system, the IDI (Integrated Data Infrastructure) [494] in ANZ. The IDI had not previously been used to investigate the ANZ dementia prevalence. Following the published article, I discuss additional information not contained in the publication, and conclude with the impact and implications of my research.

4.3 Peer-reviewed published manuscript

The following published work is reproduced under the Creative Commons CC BY licence: **Walesby KE**, Exeter DJ, Gibb S, Wood PC, Starr JM, Russ TC. Prevalence and geographical variation of dementia in New Zealand from 2012 to 2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure. *Australas J Ageing*. 2020 May 11. doi: 10.1111/ajag.12790 [3].

Prevalence and geographical variation of dementia in New Zealand from 2012 to 2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure

Katherine Elizabeth Walesby^{1,2} | Daniel John Exeter³ | Sheree Gibb⁴ | Philip Clive Wood^{5,6,7} | John Michael Starr^{1,8*} | Tom Charles Russ^{1,2,9,10,11}

¹Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK

²Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

³Epidemiology and Biostatistics, School of Population Health, University of Auckland, Auckland, New Zealand

⁴Department of Public Health, University of Otago, Wellington, New Zealand

⁵North Shore Hospital, Auckland, New Zealand

⁶Auckland Dementia Prevention Research Clinic, Auckland, New Zealand

⁷Healthy Ageing, Ministry of Health New Zealand, Wellington, New Zealand

⁸Western General Hospital, NHS Lothian, Edinburgh, UK

⁹Centre for Dementia Prevention, University of Edinburgh, Edinburgh, UK

¹⁰Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

¹¹NHS Lothian, Edinburgh, UK

Correspondence

Katherine Elizabeth Walesby, Alzheimer Scotland Dementia Research Centre, 7 George Square, University of Edinburgh, EH8 9JZ Edinburgh, UK.
Email: Katherine.walesby@ed.ac.uk

Funding information

Medical Research Council (MRC) UK, Grant/Award Number: MR/L501530/1 and MR/K02992/1; Medical Research Council Supplementary Travel Grant; Biotechnology and Biological Sciences Research Council; Alzheimer Scotland, Grant/Award Number: R44394

Abstract

Objectives: There are no national dementia epidemiological studies using New Zealand (NZ) data. NZ routinely collects health-care data within the Integrated Data Infrastructure (IDI). The study objectives were to 1) investigate late-onset dementia estimates using the IDI between 2012–2015 and compare these with 2) published estimates, and 3) variations between North and South Islands and ethnicity.

Methods: A population-based, retrospective cohort design was applied to routinely collected de-identified health/administrative IDI data. Dementia was defined by ICD-10-AM dementia codes or anti-dementia drugs.

Results: Approximately 2% of those aged ≥60 years had dementia, lower than published estimates. Dementia was higher in North Island; in 80- to 89-year-olds; among the Māori population when age-standardised, and 9% of all dementia cases had >1 dementia sub-type.

Conclusions: To our knowledge, this is the first study ascertaining dementia estimates using NZ's whole-of-population IDI data. Estimates were lower than existing

Daniel John Exeter and Sheree Gibb joint second authors.

*Deceased.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Australasian Journal on Ageing* published by John Wiley & Sons Australia, Ltd on behalf of AJA Inc

NZ estimates, for several reasons. Further work is required, including expanding IDI data sets, to develop future estimates that better reflect NZ's diverse population.

KEYWORDS

Dementia, epidemiology, New Zealand, prevalence

1 | INTRODUCTION

Dementia is an important public health concern, predicted to affect 75.62 million people worldwide in 2030.¹ Understanding country-specific burden of disease facilitates effective health-care delivery. There have been no national New Zealand (NZ) epidemiological studies of dementia.²⁻⁴ One valuable study (LiLACS NZ) investigated dementia prevalence using NZ data but not at a national level.^{5,6} LiLACS NZ uses a small population within two districts of North Island and focuses on the older age groups (≥ 80 -year-old Māori and ≥ 85 -year-old Non-Māori).^{5,6} However, estimates for national dementia prevalence have previously been produced by applying dementia prevalence rates for Australia to the estimated population of NZ^{3,4} or by using incidence rates for Alzheimer dementia from an international systematic review modelled on NZ's ageing population.² The Alzheimer Disease International (ADI) 2015 report produced regional (Australasian) dementia prevalence estimates. Therefore, the recent NZ estimates apply the Australian prevalence rates to the United Nations projections for NZ's population.⁴ Given the diverse ethnic compositions of NZ and Australia, we contend that whilst these ADI reports are useful, they are unlikely to accurately reflect the situation of NZ. Variations in dementia prevalence between and within countries are well recognised,^{7,8} including dementia mortality differences between NZ's North and South Islands.⁷

Greater information on dementia prevalence in NZ is required, and NZ's linked health and administrative population data within the Integrated Data Infrastructure (IDI)⁹ provide some data to study dementia at a population level. The IDI, a large research database containing microdata about people and households,⁹ contains mortality, medication dispensing and hospital discharge data sets necessary to ascertain dementia, using methods similar to in previous international research.^{8,10-12}

The aims of this study were to: (a) investigate NZ's late-onset dementia prevalence using the IDI for 2012 to 2015 (inclusive); (b) compare this with published NZ estimates; and (c) compare variations between the North and South Islands and by ethnicity.

2 | METHODS

A population-based retrospective cohort study was used analysing routinely collected de-identified health and

Policy Impact

This research highlights discrepancies between estimates of New Zealand (NZ) dementia prevalence and those within the Integrated Data Infrastructure, emphasising the need for a more detailed epidemiological study within NZ to provide country-specific dementia data to help guide health care and policies.

administrative data stored within NZ's IDI system. Health data comprised Primary Health Organisation; publicly funded hospital discharges (National Minimum Dataset [NMDS]); subsidised dispensing (Ministry of Health [MOH] and Pharmaceutical Management Agency [PHARMAC]); and mortality, cause of death, data (MOH). Other administrative data comprised Inland Revenue (pensions) and Accident Compensation Corporation (ACC) data sets.

2.1 | Data access, linkage and cohorts

Data were stored securely by Statistics NZ, with access granted following approvals.⁹

The IDI October 2016 version (Appendix S1) was used to create annual cohorts for 2012–2015. Our population at risk of dementia (denominator) was defined within the IDI as all individuals aged ≥ 60 years (alive or died during the year) and NZ residents who had interacted with any of the health data sets above, Inland Revenue (pensions) or ACC data sets. Our numerator therefore represented those individuals defined as having dementia in that year. We also obtained age, sex and multi-response ethnicity.^{13,14}

For each year, dementia cases (numerators) were counted using the case definition of dementia. Using SAS and SQL, the cohort was linked to health data sets (NMDS, dispensing and mortality data).

2.2 | Dementia ascertainment

Dementia was defined by any mention of ICD-10-AM (Australian modification version) dementia codes in NMDS or mortality data sets; or anti-dementia drug prescriptions in

TABLE 1 Dementia prevalence, dementia sub-types and demographics for the cohorts 2012-2015

Study population—population ≥60 years in NZ per year (denominator group, population at risk)	2012	2013	2014	2015
Total people per year ≥60 years old in NZ	891 558	916 065	945 252	966 483
Sex (% of total study population) ^a				
Male	409 173 (46%)	422 199 (46%)	436 506 (46%)	446 574 (46%)
Female	474 429 (53%)	487 983 (53%)	502 815 (53%)	513 975 (53%)
Ethnicity ^b				
Māori	59 145	62 208	65 508	68 670
Pacific	25 572	26 562	27 957	29 100
European	742 536	761 670	781 581	794 670
Asian	44 076	47 982	52 320	56 073
Middle Eastern, Latin American and African (MELAA)	2820	3042	3336	3627
Other	13 602	14 607	15 477	16 137
Age range ^a				
60-64	234 312	237 288	242 226	242 553
65-69	201 471	214 593	224 991	231 792
70-74	159 402	162 735	167 277	171 084
75-79	114 567	118 638	123 531	131 064
80-84	88 539	88 965	90 126	90 180
85-89	55 209	56 601	58 026	59 397
90-94	23 409	24 549	26 139	26 802
95+	6696	6819	7002	7680
Total people by North/South Island ^c				
North Island	637 245	660 852	683 220	699 813
South Island	230 001	237 825	244 248	248 934
<i>Dementia prevalence</i>				
All dementia cases—numerator (% of total study population)	13 677 (2%)	13 155 (1%)	13 800 (2%)	14 391 (2%)
Dementia by sub-type: (% of total dementia cases per year)				
Alzheimer dementia	2499 (18%)	2094 (16%)	2142 (16%)	2166 (15%)
Vascular dementia	1257 (9%)	1284 (10%)	1338 (10%)	1407 (10%)
Parkinson's disease dementia	447 (3%)	432 (3%)	441 (3%)	447 (3%)
Other (including unspecified dementia sub-type) dementia	10 647 (78%)	10 563 (80%)	11 166 (81%)	11 736 (82%)
More than one type of dementia	1110 (8%)	1140 (9%)	1194 (9%)	1272 (9%)
Dementia by sex				
Male (% of total study population per year)	5418 (1%)	5601 (1%)	5856 (1%)	6240 (1%)
Female (% of total study population per year)	8091 (1%)	7542 (1%)	7935 (1%)	8136 (1%)
Dementia by ethnicity: (% of total study population ethnicity per year)				
Māori	699 (1%)	741 (1%)	801 (1%)	867 (1%)
Pacific	279 (1%)	330 (1%)	351 (1%)	348 (1%)
European	12 057 (2%)	11 730 (2%)	12 225 (2%)	12 750 (2%)
Asian	321 (1%)	357 (1%)	411 (1%)	444 (1%)
MELAA	30 (1%)	27 (1%)	42 (1%)	51 (1%)
Other	36 (0.3%)	57 (0.4%)	75 (0.5%)	78 (0.5%)
Dementia by ethnicity, age and sex standardised (% of total study population ethnicity per year)				
Maori	2%	2%	2%	2%

(Continues)

TABLE 1 (Continued)

Study population—population ≥60 years in NZ per year (denominator group, population at risk)	2012	2013	2014	2015
Pacific	2%	2%	2%	2%
Euro	1%	1%	1%	1%
Asian	1%	1%	1%	1%
MELAA	2%	1%	2%	2%
Other	0.4%	1%	1%	1%
Dementia by age range: (% of total dementia cases per year)				
60-64	267 (2%)	300 (2%)	312 (2%)	294 (2%)
65-69	582 (4%)	621 (5%)	705 (5%)	756 (5%)
70-74	1257 (9%)	1326 (10%)	1404 (10%)	1443 (10%)
75-79	2175 (16%)	2205 (17%)	2340 (17%)	2553 (18%)
80-84	3264 (24%)	3219 (24%)	3255 (24%)	3375 (23%)
85-89	3486 (25%)	3348 (25%)	3447 (25%)	3510 (24%)
90-94	1860 (14%)	1683 (13%)	1869 (14%)	1899 (13%)
95+	618 (5%)	438 (3%)	465 (3%)	546 (4%)
Dementia by geographical area: (% of total dementia case per year)				
North Island	9696 (71%)	9363 (71%)	9870 (72%)	10 197 (71%)
South Island	3702 (27%)	3492 (27%)	3570 (26%)	3810 (27%)

Note: In accordance with the journal style, percentages have been rounded up to the nearest whole number. Disparities in sum of counts: Mortality data were only available for 2012 at the time of analysis and so do not contribute to dementia counts for other years. Counts for dementia sub-type represent an ICD coding and not an individual (who may have more than one sub-type), so the totals differ from the total dementia cases (which also includes counts found from dementia drugs as stated in methods). Counts for ethnicity do not all add up as total response ethnicity coding is used in NZ where an individual can belong to more than one ethnic group.¹³ Released outputs were subject to statistical disclosure control in accordance with Statistics Protocols, including suppressing counts <6, and random rounding to the base of three (RR3) to avoid identification.¹⁴

^aSex and age missing 7956 (2012), 5880 (2013), 5931 (2014), 5931 (2015).

^bEthnicity was missing for 38 265 in the total study population.

^cGeographical information on Islands missing 24 315 (2012), 17 385 (2013), 17 781 (2014), 17 736 (2015).

the dispensing data set (Appendix S2). The mortality data set was only available for 2012 at the time of analysis due to temporal lag in public releases. Formulation codes identified anti-dementia drugs (Appendix S3). Only two anticholinesterases (donepezil and rivastigmine) are funded, rivastigmine only since 2014. Dementia sub-types were classed by ICD-10-AM codes obtained from hospital or mortality data (Appendix S4). Appendix S5 refers to information on the SAS code.

2.3 | Statistical analyses

Dementia cases were calculated by year, sub-type and geographical region. Released outputs were subject to statistical disclosure control in accordance with Statistics NZ, including random rounding to base 3 (RR3), where numbers are rounded to the nearest multiple of 3, to protect confidentiality.^{15,16} For ethnic comparisons of dementia rates, we used direct age-sex standardisation to take account of differences in age and sex profile between

different ethnic groups. Age-sex standardised rates were calculated using 5-year age-sex groups and weights derived from the total (all ethnicities) denominator population described above.

2.4 | Ethics

Ethical approval was granted by The University of Auckland's Human Participant's Ethics Committee in April 2017 (reference 019125).

The study is reported according to RECORD (STROBE) guidelines.¹⁷

3 | RESULTS

3.1 | Denominator cohorts from 2012-2015

Table 1 details the study's population: 891 558; 916 065; 945 252; and 966 483 people for 2012-2015, respectively.

3.1.1 | Dementia prevalence

Dementia cases were 13 677 (2012), 13 155 (2013), 13 800 (2014) and 14 391 (2015)– 0.3% of NZ's total population in 2015 and 2% of our cohort aged ≥ 60 years. Most were 'unspecified dementia' sub-type; and around 9% had more than one dementia sub-type. Table 1 provides further details of dementia cases. The mean age for dementia ranged from 81.9 to 82.5 years (SD 7.7–7.8). Most dementia was amongst the 80–89 age groups (47%–49% of all dementia cases). However, in 2012 the ≥ 95 age group had a slightly higher proportion (5%), potentially reflecting the additional data set (mortality) present for this year.

Europeans were the largest ethnic group contributing most to the overall dementia counts (88%–89% of dementia cases). Māori with dementia contributed 5%–6% of dementia cases, whilst the Asian ethnic group contributed 2%–3% overall. This roughly reflects the prevalence of each ethnic group within the study population. Interestingly, when dementia cases were separated by ethnicity and calculated as a proportion of the ethnic group overall, the numbers differed less substantially (Europeans 2% vs 1% of Māori, Table 1). However, following age and sex standardisation, the ethnic variation changed, with dementia cases higher for Māori than Non-Māori (2% vs 1%, Table 1).

Figure 1 identifies dementia ascertainment overlap between IDI data sources in 2012. From individual data sets, most cases were ascertained from ICD-10-AM codes in NMDS ($n = 5751$), followed by dispensing data ($n = 4110$). Of 33 030 people who died in 2012, 1167 (4%) had dementia mentioned on death certificates (Figure 1). The remaining dementia cases were identifiable in more than one data set. However, only 120 dementia cases were found in all three data sets. In 2012, only 10% ($n = 1335$) were found in both NMDS and dispensing data sets and 8% ($n = 1128$) were found in NMDS and mortality data. This highlights that multiple data sets are beneficial for ascertaining dementia.

Donepezil prescriptions contributed 5634 (41%), 6474 (49%), 6843 (50%) and 7122 (49%) of dementia cases in 2012–2015, respectively. Rivastigmine prescriptions only contributed 57 (0.4%) and 264 (2%) dementia cases in 2014–2015, respectively (it was only publically available in NZ from 2014).

Given the population distribution, dementia was higher in North than South Island. The overall percentage of total dementia remained relatively consistent across the years in both Islands (see Table 1).

4 | DISCUSSION

Our study's purpose was to demonstrate the potential to estimate dementia prevalence using routinely collected data within NZ's national repository, the IDI. To our knowledge,

this is the first to do this. Our reported dementia estimates were as follows: *lower* than those previously published for NZ at 48 182 and 62 287 in 2011 and 2016, respectively^{3,4}; *higher* amongst Māori when results are age and sex standardised; *higher* in the North than the South Island; and 9% had more than one dementia sub-type.

However, our estimates should be interpreted cautiously and might be considered as 'lower bound' estimates. Several explanations for these variations should be considered for future work. Currently, NZ estimates use countries' data that may not be fully representative to NZ given it has previously been shown that significant intra-country variability in dementia rates exists.^{7,8} Alternatively, secondary care and mortality data can underestimate dementia counts, a well-established issue with big data. Not everyone with dementia receives a diagnosis, is hospitalised, is prescribed anti-dementia drugs or has dementia recorded on the death certificate. Many people with dementia are managed in primary care, and without these records being part of the IDI, or indeed a national requirement to collect this diagnostic information, the estimates may not include these people.

Unspecified dementia was the commonest sub-type. Around 9% of individuals had more than one dementia diagnosis/sub-type (Table 1). Without an agreed ICD coding for mixed dementia, this is not unexpected. Whilst Alzheimer's remains the commonest diagnosed dementia, recent work identified under-reporting of cerebrovascular pathology at postmortem.¹⁸ This could be important for NZ if cerebrovascular pathology is higher amongst some ethnic groups.⁵ However, inter-ethnic sub-type variations in NZ may be minor.¹⁹

Dementia counts varied between the North and South Islands. Possible explanations include age and ethnicity differences between North and South Islands; North Island substantially higher population (3/4 of NZ's population lives in North Island); or the accessibility to memory clinics to obtain a diagnosis (only seven of the 20 District Health Boards in NZ have at least one memory clinic with 5 of these in North Island)²⁰ Additionally, Non-Māori living in urban locations are more likely to be prescribed anti-dementia drugs.²¹ Further work is required to understand factors influencing geographical variation in dementia.

Recording of dementia may be lower in ethnic minorities.¹⁰ Despite representing 15% of NZ's population, Māori had 6% of total dementia cases in 2015. However, following age and sex standardisation, Māori dementia was approximately 50% higher than Europeans'. Recent NZ studies highlight possible ethnic differences in diagnosis-seeking and earlier onset of dementia,^{19,22,23} potentially also influenced by lower life expectancy.^{19,22,23} However, the LiLACS NZ study showed no significant difference in dementia rates between Māori and Non-Māori at an older age (>80 years) within a smaller cohort (937 people)⁵. Nonetheless, the attitudes to dementia, including the stigma amongst indigenous

people, are importantly highlighted in a recent ADI report²⁴, with a recent study discussing the complexities surrounding diagnosing *mate wareware* (dementia) in Māori communities.²⁵ Further investigation, ideally within a national epidemiological study of dementia, is warranted.

Population-based studies using big data have several limitations. Of relevance is low diagnostic coverage for dementia²⁶ or the 'diagnosis gap', which comprises three important and inter-related dimensions: patients who have dementia but do not have it recorded as a diagnosis; those who have dementia but do not interact with health/social services; and patients who have cognitive impairment that impacts on daily function but may not receive a dementia-related diagnosis. In the UK, specific population studies have been used to estimate the 'diagnosis gap' and then extrapolated to the whole population to better inform health-care delivery.²⁷

Further limitations include the types and reliability of data sets for ascertaining dementia in NZ's multi-ethnic population. Neither memory clinic nor primary care consultation data are available within the IDI. Primary care consultation data inferring cognitive impairment and dementia would be a valuable addition to the IDI, as would memory clinic outpatient data. Indeed, since our study, the IDI now incorporates the NZ InterRAI database (an extensive resource for understanding cognition, dementia and care needs of NZ's older people).²⁸ This addition, along with including other data sets into the IDI, will allow validation of data sets for dementia coding and ascertainment specific to NZ and triangulation of dementia prevalence results to facilitate the delivery of NZ's Framework for Dementia Care.²⁹

Wilkinson's recent systematic review showed dementia coding and positive predictive values (PPVs) varied widely,

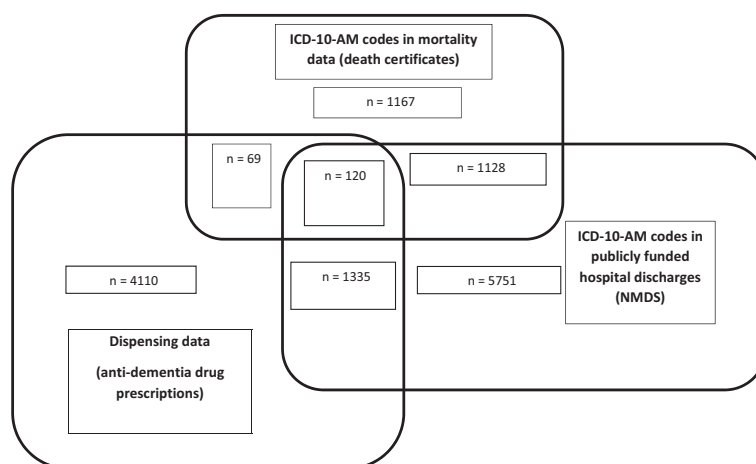
with <50% of dementia cases identified from routinely collected data and PPVs for dementia diagnoses ranging from 33% to 100%.³⁰ This is comparable with other studies,^{10-12,31-33} one highlighting only 53% of known dementia patients had dementia recorded on their hospital discharge.³⁴ However, some studies achieved PPVs of >80%-90% within routinely collected data, meaning that those with a dementia diagnosis recorded were likely to have the diagnosis of dementia.³⁰

Within secondary care data, dementia ascertainment had higher sensitivity if hospital admissions and mortality data were combined, and lower sensitivity when only primary ICD codes were selected from death certificates.³⁰ We mitigated for this by including ICD codes in any position on death certificates and combining hospital admissions and mortality data. Some recent studies have shown comparability of routinely collected hospital and primary care data for dementia ascertainment,^{11,12} highlighting good sensitivity and specificity compared with memory clinic data.¹⁰ Whilst there is no 'gold-standard' data source, we attempted to overcome some of these limitations by including multiple data sets, as recommended by recent systematic reviews.^{30,35}

It is also important to recognise that routinely collected data will vary with time. For instance, rates may rise due to ageing populations, improved diagnosis, access to additional anti-dementia drugs (eg Rivastigmine in NZ in 2014) and awareness of dementia.

5 | CONCLUSIONS

This study shows it is possible to ascertain dementia counts within the IDI and that these differ from current estimates.



Note: Released outputs were subject to statistical disclosure control in accordance with Statistics NZ protocols. The Integrated Data Infrastructure data sets are updated at different time points—the mortality data set was only available for 2012 at the time of analysis due to temporal lag in its public releases.

FIGURE 1 Venn diagram of dementia ascertainment for dementia cases for 2012

Traditional 'gold-standard' country-wide epidemiological studies are costly and labour-intensive, but the IDI could provide a future tool to enhance understanding of dementia prevalence in NZ at a national level. Our research highlights the need for further work including a national epidemiological study of dementia using the InterRAI within the IDI, and adding further relevant data sets (memory clinics and primary care). This will also allow validation for dementia recording accuracy in NZ's routinely collected data sets, improve dementia ascertainment; and understand the 'diagnosis gap' specific to NZ, and ultimately better inform health-care delivery.

ACKNOWLEDGEMENTS

KEW is supported by clinical research fellowship from Alzheimer Scotland and the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-council Lifelong Health and Wellbeing initiative (MR/L501530/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. KEW, JMS and TCR are members of the Alzheimer Scotland Dementia Research Centre funded by Alzheimer Scotland. KEW additionally received a PhD studentship travel grant from the MRC in 2016 for a preliminary visit to New Zealand for planning the research and an additional grant from Alzheimer Scotland to undertake the research in 2017. The funders played no part in the design or conduct of this manuscript. Thanks to the IDI team at Statistics New Zealand for their valuable input, assistance with the study and use of data, in particular thanks to Simon McBeth and Stephen Challands. Additionally, we would like to thank Dr Jinfeng Zhao for the original VARIANZ population SAS code (co-written) with Dr Sheree Gibb that this study's SAS code was based on. Lastly, we would like to thank Professor John Starr (co-author) for his huge lifetime contribution to dementia research worldwide. We are deeply saddened by his sudden unexpected death on 9 December 2018 and miss our colleague greatly.

CONFLICTS OF INTEREST

No conflicts of interest declared.

DISCLAIMER STATEMENT

Access to the data presented was managed by Statistics New Zealand under strict microdata access protocols and in accordance with the security and confidentiality provisions of the Statistic Act 1975. Our findings are not Official Statistics. The opinions, findings, recommendations and conclusions expressed are those of the researchers, not Statistics NZ, or the Universities of Edinburgh or Auckland.

ORCID

Katherine Elizabeth Walesby  <https://orcid.org/0000-0003-3401-3793>

Daniel John Exeter  <https://orcid.org/0000-0003-1061-5925>

Tom Charles Russ  <https://orcid.org/0000-0001-9797-2188>

REFERENCES

1. Prince M, Guerchet M, Prina M. *Alzheimer's Disease International: Policy Brief for Heads of Government, The Global Impact of Dementia 2013-2050*. London: Alzheimer's Disease International; 2013.
2. Tobias M, Yeh LC, Johnson E. Burden of Alzheimer's disease: population-based estimates and projections for New Zealand, 2006–2031. *Aust N Z J Psychiatry*. 2008;42(9):828–836.
3. Deloitte Access Economics. *Updated Dementia Economic Impact Report, 2011, New Zealand, Alzheimers New Zealand*. 2012:i-32. http://www.alzheimers.org.nz/getattachment/News-Info/New-Zealand-data/Updated_Dementia_Economic_Impact_Report_2012_New_Zealand.pdf/. Accessed March 5, 2020.
4. Deloitte and Alzheimers New Zealand. *Dementia Economic Impact Report, 2016*. 2016: 1–78. <https://www.alzheimers.org.nz/getmedia/79f7fd09-93fe-43b0-a837-771027bb23c0/Economic-Impacts-of-Dementia-2017.pdf/>. Accessed March 5, 2020.
5. Kerse N, Lapsley H, Moyes S, Zawaly K, Hayman K. *LiLACS NZ 2017. Dementia: supplementary findings from LiLACS NZ for Section Five, 'Service Use and Common Health Conditions' in the report 'Health, Independence and Caregiving in Advanced Age'*. Auckland. In: LiLACS NZ, School of Population Health, ed. The University of Auckland; 2017. <https://www.fmhs.auckland.ac.nz/en/faculty/lilacs/research/publications.html>. Accessed March 5, 2020.
6. Teh R, Kerse N, Kepa M et al. Self-rated health, health-related behaviours and medical conditions of Maori and non-Maori in advanced age: LiLACS NZ. *N Z Med J*. 2014;127(1397):13–29.
7. Russ TC, Murianni L, Icaza G, Slachevsky A, Starr JM. Geographical variation in dementia mortality in Italy, New Zealand, and Chile: the impact of latitude, vitamin D, and air pollution. *Dement Geriatr Cogn Disord*. 2016;42(1–2):31–41.
8. Russ TC, Gatz M, Pedersen NL et al. Geographical variation in dementia: examining the role of environmental factors in Sweden and Scotland. *Epidemiology*. 2015;26(2):263–270.
9. Statistics New Zealand. *Integrated Data Infrastructure*; 2018. <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>. Accessed March 5, 2020.
10. Sommerlad A, Perera G, Singh-Manoux A, Lewis G, Stewart R, Livingston G. Accuracy of general hospital dementia diagnoses in England: Sensitivity, specificity, and predictors of diagnostic accuracy 2008–2016. *Alzheimers Dement*. 2018;14(7):933–943.
11. Brown A, Kirichek O, Balkwill A et al. Comparison of dementia recorded in routinely collected hospital admission data in England with dementia recorded in primary care. *Emerg Themes Epidemiol*. 2016;13:11.
12. Pujades-Rodriguez M, Assi V, Gonzalez-Izquierdo A et al. The diagnosis, burden and prognosis of dementia: a record-linkage cohort study in England. *PLoS ONE*. 2018;13(6):e0199026.

13. Ministry of Health New Zealand. *HISO 10001:2017 Ethnicity Data Protocols*. Wellington; 2017. <https://www.health.govt.nz/publication/hiso-100012017-ethnicity-data-protocols>. Accessed March 5, 2020.
14. Teng A, Gibb S, Sporle A. *Ethnicity in the IDI*; 2017 [cited 2019 18th June]. <https://vhin.co.nz/guides/ethnicity-and-the-idi/>. Accessed March 5, 2020.
15. Statistics New Zealand. *Microdata Output Guide*; 2016. <http://archive.stats.govt.nz/~media/Statistics/services/microdata-access/data-lab/microdata-output-guide-2016.pdf>. Accessed March 5, 2020.
16. Statistics New Zealand. *Standards*; 2019. http://archive.stats.govt.nz/browse_for_stats/Maps_and_geography/Geographic-areas/urban-ruralprofile/standards.aspx. Accessed March 5, 2020.
17. Benchimol EI, Smeeth L, Guttman A et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.
18. Grandal Leiros B, Pérez Méndez LI, Zelaya Huerta MV et al. Prevalence and concordance between the clinical and the post-mortem diagnosis of dementia in a psychogeriatric clinic. *Neurologia*. 2018;33(1):13-17.
19. Cullum S, Mullin K, Zeng I et al. Do community-dwelling Maori and Pacific peoples present with dementia at a younger age and at a later stage compared with NZ Europeans? *Int J Geriatr Psychiatry*. 2018;33(8):1098-1104.
20. Stone C, Copeland B, Collier C, Cheung G. Memory clinic survey in New Zealand: a second look. *Australas Psychiatry*. 2019;27(5):486-490.
21. Norris P, Horsburgh S, Lovelock K et al. Medicalisation or under-treatment? Psychotropic medication use by elderly people in New Zealand. *Health Sociol Rev*. 2011;20(2):202-218.
22. Payman V, Yates S, Cullum S. Early onset dementia in New Zealand Pacific boxers: a case series. *N Z Med J*. 2018;131(1474):20-26.
23. Dyall L. Dementia: continuation of health and ethnic inequalities in New Zealand. *N Z Med J*. 2014;127(1389):68-80.
24. Alzheimer's Disease International. *Alzheimer's Disease International World Alzheimer Report 2019: Attitudes to dementia*. London: Alzheimer's Disease International (ADI); 2019. <https://www.alz.co.uk/research/WorldAlzheimerReport2019.pdf>. Accessed March 5, 2020.
25. Dudley M, Menzies O, Elder H, Nathan L, Garrett N, Wilson D. Mate wareware : Understanding 'dementia' from a Māori perspective. *N Z Med J*. 2019;132(1503):66-74.
26. Prince M, Comas-Herrera A, Knapp M, Guerchet M, Karagiannidou M. *World Alzheimer Report 2016: Improving healthcare for people living with dementia. Coverage, Quality and Costs now and in the future*. London: Alzheimer's Disease International (ADI); 2016. <https://www.alz.co.uk/research/WorldAlzheimerReport2016.pdf>. Accessed March 5, 2020.
27. Dementia Intelligence Network. *Dementia Partnerships*. <https://dementiapartnerships.com/resource/the-dementia-profile/>. Accessed March 5, 2020.
28. Technical Advisory Services (TAS) Kahui Tuitui Tangata and interRAI New Zealand. *National interRAI Data Analysis Annual Report 2015/16*; 2017. <https://tas.health.nz/assets/interRAI/2015-16-National-interRAI-Data-Analysis-Annual-Report.pdf>. Accessed March 5, 2020.
29. Ministry of Health New Zealand. *New Zealand Framework for Dementia Care*. Wellington: Ministry of Health New Zealand; 2013:i-49.
30. Wilkinson T, Ly A, Schnier C et al. Identifying dementia cases with routinely collected health data: a systematic review. *Alzheimers Dement*. 2018;14(8):1038-1051.
31. Waller M, Mishra GD, Dobson AJ. Estimating the prevalence of dementia using multiple linked administrative health records and capture-recapture methodology. *Emerg Themes Epidemiol*. 2017;14:3.
32. Francesconi P, Gini R, Roti L, Bartolacci S, Corsi A, Buiatti E. The Tuscany experimental registry for Alzheimer's disease and other dementias: how many demented people does it capture? *Aging Clin Exp Res*. 2007;19(5):390-393.
33. van de Vorst IE, Vaartjes I, Sinneker LF, Beks LJM, Bots ML, Koek HL. The validity of national hospital discharge register data on dementia: a comparative analysis using clinical data from a university medical centre. *Neth J Med*. 2015;73(2):69-75.
34. Russ TC, Parra MA, Lim AE, Law E, Connelly PJ, Starr JM. Prediction of general hospital admission in people with dementia: cohort study. *Br J Psychiatry*. 2015;206:153-159.
35. Sibbett RA, Russ TC, Deary IJ, Starr JM. Dementia ascertainment using existing data in UK longitudinal and cohort studies: a systematic review of methodology. *BMC Psychiatry*. 2017;17(1):239.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Walesby KE, Exeter DJ, Gibb S, Wood PC, Starr JM, Russ TC. Prevalence and geographical variation of dementia in New Zealand from 2012 to 2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure. *Australas J Ageing*. 2020;39:297–304. <https://doi.org/10.1111/ajag.12790>

4.3.1 Supplementary manuscript information – not peer reviewed

This supplementary information for the manuscript is available online separately (as Appendices 1-5) to the above published peer-reviewed article. This supplementary information is not peer reviewed, but submitted to the journal along with the manuscript. I discuss this supplementary information further in this Chapter. I include it here (as submitted to the journal) for clarity in this Chapter.

Appendix 1

Additional Methods information on the datasets used within the IDI system

The datasets within the IDI system are added and can lag behind the real-time generation of these datasets in clinical practice.

Of the datasets available in the October 2016 refresh of the IDI, we looked at 2012-2015. As listed below the datasets were not all available until December 2015 and a summary of their updated end date at the time of the analysis is listed below:

- Mortality data: was available to the end of December 2012 for cause of death information
- Primary care registration was available until June 2015
- Publicly funded hospital discharges was available until December 2015;
- Pharmaceutical data was available until March 2016;
- ACC data was available until July 2015

For ethnicity, we used national ethnicity coding protocols within the IDI based on multi-source ethnicity ranking rather than prioritised ethnicity ranking which is commonly used in health research. Further information can be found on this method of classifying ethnicity.[13] This is preferred as it provides a single ethnic profile for each individual within the IDI and is recommended by Statistics NZ as it has been found to have best agreement with the census and health datasets. However, a limitation of it is that it might underrepresent those individuals who associate with multiple ethnicities

The mortality dataset has a temporal lag for public releases due to the wait for completed coronial hearings. Therefore the mortality dataset is not as up to date as the other datasets in the IDI due to these delays. At the time of the analysis, the mortality dataset was only available for 2012.

Appendix 2

Dementia Ascertainment from ICD 10-AM codes

	ICD10-AM Code
Diagnosis of Dementia	F00
	F000
	F001
	F002
	F009
	F01
	F010
	F011
	F012
	F013
	F018
	F019
	F02
	F020
	F021
	F022
	F023
	F024
	F028
	F03
	F051
	F107
	F137
	F187
	F197
	G30
	G300
	G301
	G308
	G309
	G310
	G311
	G313

Appendix 3

Dementia Ascertainment from Prescription of Dementia Specific Medications available in New Zealand

Anti-Dementia specific drug	Formulation ID in PHARMAC	Probable Diagnosis of Dementia	Years publicly funded/available in New Zealand
Donepezil	392326	YES	All cohorts (2012-2015)
Donepezil	392325	YES	
Rivastigmine	403725	YES	Cohorts 2014-2015 only
Rivastigmine	403726	YES	

Appendix 4

List of dementia sub-types (ICD-10-AM codes)

Dementia Sub-type	ICD 10-AM Codes included in sub-type
Alzheimer Disease	F00 F000 F001 F002 F009 G30 G300 G301 G308 G309
Vascular Dementia	F01 F010 F011 F012 F013 F018 F019
Parkinson's Disease Dementia	F023 G313
Other Dementia (including Fronto-Temporal Dementia and Unspecified Dementia)	F02 F03 F020 F021 F022 F024 F028 F051 F107 F137 F187 F197 G310 G311

For outputting the data, where counts in certain subtypes of dementia were too small these were merged merging with 'Other Dementia' sub-type

Appendix 5 – Accessibility, raw data and programming code

- **Accessibility and raw data**

The authors do not have permission to grant access to the raw Individual level data. Access to this information is granted via Statistics New Zealand and only available within their secure data labs following their application procedures.

- **Programming code**

The SAS code is available from the authors on request. Two SAS codes were created:

- *SAS programming code for creating the denominator group (population at risk)*
- *SAS programming code for creating the numerator group (those with dementia case definitions) referred to as DEMGEO (Dementia Geography) population in the code*

This was the first whole-of-population estimate of late-onset dementia in ANZ using routine datasets. The benefits of undertaking this project are threefold. First, this manuscript generated vital new knowledge on the scale of dementia specific to ANZ in addition to the geographical distribution of the population diagnosed with dementia. Second, this research provided a nationally representative baseline to be the foundation for future investigations that aim to manage and support those with dementia in ANZ. Third, this research may inform health policy on the healthcare resources, services and delivery. Experience from the UK [152, 527, 438] suggests that dementia prevalence studies, using existing datasets, are feasible and provide valuable data to improve healthcare planning and dementia prevention.

The remainder of this chapter discusses the background to planning and developing the ANZ-based research (including discussing why my research plans and question changed following the discovery of routinely collected national healthcare data in ANZ). I also outline the importance of data reflecting the population it represents, why ANZ was chosen to investigate geographical variation of dementia, and discuss data available in ANZ that I considered for my research.

4.4 Importance of representative data

There is a global movement in which health data are expected to be available widely and be used to save lives, with an associated hashtag #datasaveslives [569]. However, for this to be meaningful, the data has to be representative of the underlying population. I now discuss the importance of using country-specific data, in the context of ANZ, which is a country of many ethnicities, understanding the dementia prevalence between the different ethnicities to improve how dementia services, treatment and care are received. ANZ is not alone in this respect and reinforces the benefit of being able to access routinely collected data at a population level to inform estimates and services for dementia. Extrapolation

of other countries' data may not be representative, particularly in ANZ with its important indigenous and multi-ethnic population.

Ethnicity within routinely collected data

Ethnicity is considered a multidimensional social construct [570, 571]. Ethnicity can be defined as: *a social construct used to characterise a group of people who share commonalities in ancestry, history, culture, traditions, values, and beliefs* [572, 571, 570, 573]. There are three overlapping dimensions: ancestry, self-identification, and observer-identification [574].

Ethnicity originated from the biological concept of race, which has been documented to have gained popularity in the 18th century due to attempts to justify colonisation and slavery [572, 574, 575, 576]. The traditional concept of race organised humans hierarchically based on innate physical characteristics stemming supposedly from genetic composition (e.g., skin colour), and as highlighted was supported by racial (pseudo-) science at the time [572, 570, 576]. Subsequent developments in genetic research highlighted that the biological premise for race was limited and incorrect, as genetic between-group differences were found [572]. Following World War II, race began to evolve into the social construct now called “ethnicity” [570].

Historically, due to the biological concept of race, an individual's race (and ethnicity) was typically classified by colonised states to control and dominate. This also resulted, most notably in the United States of America, in limiting multi-racial people from being allowed to self-identify with more than one race [572, 574, 575, 576]. A consequence of this was forced “hypodescent” amongst multi-racial children (the automatic assignment of children of multi-racial parents to what society deemed the subordinate group) to prevent assimilation, and increase slavery [572, 574, 575].

Whilst society has moved forwards from these awful examples, it is well known

that there remains many inequalities in race and ethnicities both at an individual, and institutional, level. Understanding ethnicities in the country context is really important. Ethnicity is of importance in current research as it is essential to ethnic group recognition, and also for equity and anti-discrimination agendas [572, 577, 578, 579, 580, 581, 582] Translating results from one country to another is not necessarily valid for many cultural and ethnic reasons. It is also important to understand that countries may collect information on an individual's multiple ethnic identifications, but classify these into a smaller number for statistical outputs of population prevalences. Understanding this, within routinely collected data, and how this may vary between countries is important. Globally, ethnicity is typically collected via self-identification using a “check-all-that-apply” approach [571] so individuals can appropriately highlight their ethnicity or multiple ethnicities they identify with.

In ANZ, prior classification for Māori, the Indigenous Peoples of the country, was restricted to those who reported at least “half” Māori descent [583, 570, 577, 584, 585]. This did not allow an accurate reflection of an individual's ethnicity/ethnicities. Additionally, in the past, as research in ANZ did not always appropriately address ethnicity, this led to further exclusion. This is now being carefully addressed.

No universal “best” ethnic classification method exists. Experts believe that the most appropriate ethnic classification method can depend on the research question, and sociocultural context [583, 578, 570, 577, 576, 585]. For multivariable statistical analysis, researchers typically select classification methods that output multiple ethnic identifications into a smaller number of mutually exclusive categories [577, 579, 570, 586, 585]. Two common methods are: 1) administrative-prioritisation, and 2) self-prioritisation [570, 587]. Administrative-prioritisation refers to a coding scheme which prioritises multiple ethnic responses into a single ethnic group. Self-prioritisation is obtained through a question survey which asks multi-ethnic participants to self-select a “main” ethnic group. Previous research

indicates considerable variations in how multi-ethnic participants are classified according to these two methods [578, 577, 573, 588, 586].

ANZ has an important approach to ethnicities when capturing data, which was different to my past experience of understanding routinely collected health data in Scotland. Patients interacting with the healthcare system state the ethnic group(s) with which they identify, following national healthcare data collection protocols [587, 577, 578]. This is the self-prioritisation method. Self-identification of ethnicity in ANZ has been shown to lead to different output types (total response, sole/combination, prioritised) [570, 577, 578].

Obtaining an accurate reflection of an individual's ethnicity from routinely collected health data is complex. The ANZ Ethnicity protocols allowed patients to self-identify with multiple different ethnic groups at the time of my research, but further changes were incorporated in 2021 [587]. The changes in 2021 resulted in ANZ moving to a 'total response' in relation to an individual's ethnicity [587]. This means *"that we recognise all aspects of someone's ethnicity and allows us to represent the full diversity of communities, whānau, and people in Aotearoa. For example, if someone identifies as Māori and Tongan, they will be reported as both Māori and Tongan"* [587].

For the purposes of health data analyses, in ANZ prioritisation ethnicity is ordered into these groups: Māori, Pacific, Asian, Middle Eastern Latin American & African, and NZ European/Others [585]. European New Zealanders are the largest (62.1%) ethnic group in the recent 2023 Census (sometimes referred to by the Māori name Pākehā meaning white person) [589]. Māori are the second largest ethnic group (17.8%). The Māori population varied between regions of ANZ, but all regions had an increase from 2018-2023 Census [589]. There are also many people from other ethnicities living in ANZ, and, as highlighted, many identify with more than one ethnicity. Auckland (ANZ largest city and located in the North Island) has the highest proportion of people of Asian (31.3% compared

with 17.3% nationally), and Pacific ethnicities (16.6% with 8.9% nationally) [589]. Gisborne (North-east ANZ) had the largest population of Māori [589].

When considering my results from the published manuscript Section 4.3 (Peer-reviewed published manuscript) comparing North and South Island, it is important to recognise that population densities are different, and also that variations in ethnicities (compared with the national percentages) can differ by areas. A direct comparison between North and South Island need careful interpretation in view of this.

ANZ population demographics have changed over time. There have been more people from India, China, and the South Pacific migrating to the country in recent years. Amongst the Pākehā there are strong genetic links with Scotland, particularly in South Island (Dunedin is Gaelic for Edinburgh). ANZ is a rich multi-ethnic and multi-cultural country with an important Indigenous population. Understanding ethnicity within routinely collected data is important in view of this: highlighting that other countries' data on dementia rates may not be appropriate for the ANZ context, and that variations geographically in dementia prevalence needs to be considered holistically.

Dementia: cultural differences for this term

In Māori, the word dementia translates as mate wareware. Mate means 'sickness/disease/illness/unwell', and wareware means 'becoming forgetful' [590, 591]. Within Māori culture, the impacts of dementia on whānau (extended family), and how this may impact on seeking or receiving a diagnosis of dementia are important considerations. The findings of a report in 2019 undertaking kaupapa Māori research (Māori informed research) of 223 kaumātua (Māori elders), and seventeen focus groups across ANZ found key themes to understand 'dementia' care and management from a Māori perspective [591]. The report highlighted Māori understandings of the cause of mate wareware, and key aspects important in the care-giving for whānau living with mate wareware. These included the role

of aroha (love, compassion), and manaakitanga (hospitality, kindness, generosity, support and caring) [591].

There has been increasing kaupapa Māori research (studies of Māori participants conducted by Māori researchers) recognising the cultural impacts of dementia diagnosis, and management, on Māori in ANZ. This is an important, and growing development, since my research in ANZ. Barriers to dementia diagnoses and management will also affect estimates of dementia prevalence in ANZ as the results may be cautious estimates of the ‘true’ prevalence of dementia due to unrecognised barriers for people in seeking a diagnosis. This includes systems that do not incorporate Indigenous Peoples’ beliefs in relation to health and disease.

In Australia, a neighbouring country with an indigenous population, researchers investigating dementia incidence and prevalence amongst indigenous and non-indigenous populations demonstrated that their estimates were higher than the current national estimates [400]. That study also demonstrated that undetected cases were over-represented in the indigenous population, despite only being 30% of the study area’s population [400]. The results from Li et al. 2024 amplified previous studies that had demonstrated the urgent need to use representative country specific data when establishing national estimates for dementia prevalence [400]. As prior publications demonstrate [592, 593], there are concerns about using previously-published prevalence estimates (using non ANZ data) for modelling dementia prevalence for ANZ.

In summary, dementia must be considered in the cultural, and ethnic context of the country that research is being undertaken. To date, no research has demonstrated the extent to which the dementia prevalence for indigenous Aboriginal Australians correlates with the ANZ Māori population. At the time of writing the published manuscript in this Chapter, no representative surveys of dementia at a population level across ANZ among the Māori population had been conducted. More recent developments in ANZ, since the initial publication of my

paper in 2020, will be discussed in Section 4.10 ([Additional discussion and future work](#)).

4.5 Why investigate the geographical variation of dementia in ANZ?

A decision to investigate geographical variation of dementia between the north and south of a country was informed by previous research that demonstrated the unexplained variations between north and south of countries in the Northern Hemisphere [350, 351]. One study found a higher dementia prevalence in the north of two countries (Scotland and Sweden) [350], while another study found the reverse of this in one Southern Hemisphere country [351]. As discussed in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#) several studies have shown a north/south variation in their dementia rates. These includes studies in Chile, Finland, Italy, ANZ (death certificates only), Scotland and Sweden [351, 350, 594].

In the studies in the Northern Hemisphere, the dementia risk was generally low in the south of the countries and increasing further north. In comparison, in the study including Chile and ANZ, the standardised mortality ratios (SMRs) for deaths where dementia was recorded showed a possible inverse pattern in women in ANZ with rates higher in the South Island than the North Island, although not showing the same variation in Chile. The study by Russ et al. 2016 in ANZ [351] only used death certificates, so caution is advised when interpreting these results. These results may reflect areas in which care homes may cluster, given dementia is linked with clinical frailty, and an increased reason for admission to a care home.

Given this unexplained and possible variation in dementia rates, the question arose of whether latitude contributed to this variation of dementia between the north and south of various countries. A possible underlying reason for geographi-

cal variation of dementia in relation to latitude may involve Vitamin D, which has been previously proposed as a possible reason for variations of dementia. Vitamin D levels vary with latitude. If true, then dementia prevalence results should be reversed between Northern and Southern Hemisphere.

A number of articles have considered the role of Vitamin D and dementia, including some more recent publications since my data collection/analysis in ANZ. The Rotterdam Study of 6220 people demonstrated at baseline that lower Vitamin D concentrations were non-significantly associated with higher prevalence of all cause dementias (adjusted OR, per SD decrease 1.20, 95% CI 0.95; 1.52) [595]. When following up patients who later developed dementia, lower Vitamin D concentrations were associated with higher risk of all cause dementia (adjusted HR, per SD decrease 1.11, 95% CI 1.02; 1.20) and also Alzheimer's Dementia (adjusted HR: 1.12, 95%CI 1.03; 1.24) [595].

A previous systematic review investigating Vitamin D deficiency, cognitive impairment and dementia showed an association between cognitive impairment and Vitamin D deficiency, although there are methodological limitations in many studies investigating Vitamin D and dementia risk such as difficulties in establishing causation and confounding factors. For instance low Vitamin D levels may reflect more severe dementia or frailty with less ability to mobilise outside for sunlight absorption of Vitamin D i.e. a result of dementia rather than a cause. A more recent systematic review investigated Vitamin D supplementation and dementia showed no conclusive benefit in Vitamin D supplementation on cognition [596].

Moreover, fully understanding the role of Vitamin D across a life time, or in key points of a life time, and how Vitamin D may or may not impact on cognition and dementia is difficult. Sunlight exposure, linked with latitude, is only one aspect influencing Vitamin D levels. Therefore, the association between latitude, Vitamin D and dementia is complex due to the changes over a person's lifetime

in exposure and many other factors. To date there is no definitive causal link between dementia and lifetime Vitamin D exposure. This is in part due to the difficulty of protecting observational studies investigating Vitamin D as a risk factor for dementia from potential reverse causation for a late onset disease such as dementia (i.e. does decreased Vitamin D precede dementia onset) [597].

Work using Mendelian Randomisation to investigate the role of Vitamin D [in the form of 25-hydroxyvitamin D (25OHD) which is more stable], and dementia risk using single nucleotide polymorphisms (SNPs) supported the role of 25OHD as a causal risk factor for dementia. This is important because SNPs are assigned before disease onset and therefore reduce this reverse causation, and confounding [597]. For the purpose of my initial research plan, the original aim (which was honed after my scoping visit in 2016) was to first better understand the geographical variations of dementia prevalence across ANZ and whether this confirmed previous research suggesting higher numbers of people with dementia in Southern versus Northern parts of a country in the Southern Hemisphere.

As previously discussed in Section 4.5 ([Why investigate the geographical variation of dementia in ANZ?](#)), ANZ was a suitable country in the Southern Hemisphere with a similar national healthcare system to Scotland and geographically well spread across latitudes [598, 599] to enable the investigation of a north/south variation in dementia. Nevertheless, it is worth noting that ANZ was not initially considered due to knowledge of its routinely collected population level healthcare data. I was unaware of this valuable resource prior to my scoping visit to ANZ in 2016. Instead, I found this resource (the Integrated Data Infrastructure, IDI, [494, 600]) through discussions at ANZ Ministry of Health in Wellington.

Once I was aware of the IDI I was interested from a research perspective if it had been used for research to investigate dementia prevalence. This interest had developed after I appreciated that ANZ dementia prevalence estimates were calculated from non ANZ data and extrapolated to ANZ. After establishing that the

IDI has not previously been used for this purpose, I changed my research direction to use this valuable large data resource to investigate my research question.

4.6 Data available in ANZ for epidemiological research

In this section I discuss some of the data in ANZ that I researched prior to my initial scoping visit in 2016, during which I spoke with key researchers to understand more about the datasets and whether it would be appropriate for my research question. Finally, I discuss the Integrated Data Infrastructure (IDI) which I discovered during the scoping visit and subsequently influenced how I undertook my research question. I start by addressing what longitudinal datasets and cohort studies are available in ANZ as of 2016.

4.6.1 Longitudinal datasets and cohorts in ANZ

In preparation for the scoping visit to ANZ in 2016 I identified a number of cohorts that may have been suitable for my original research question: *‘to better understand the geographical variations of dementia prevalence across ANZ’*, Table 4.1. One cohort was nationally collected and the others were either in North or South Island. Not all of the cohorts focused on older people therefore, such as the childhood cohort Growing up in New Zealand, GUiNZ [601], so were not appropriate to consider for dementia.

It also became clear that none of these existing traditional longitudinal cohort studies were suitable to investigate my research question. Some of the cohorts (Growing up in New Zealand study, GUiNZ, and the New Zealand Health, Work and Retirement, HWR, study) [601, 602] were too young and therefore dementia ascertainment would not be appropriate. Whilst the cohorts in North and South Island were detailed cohorts, none covered the whole of the Island, the age groups of interest, nor the ethnic composition representative of ANZ.

Study	Data owner	Areas	Time period	Age range	Pop. size	Description
The New Zealand Health, Work and Retirement Study (NZHWR) [602]	Massey University	All New Zealand	2008-	50-84yrs from 2010	4,000	Longitudinal cohort. Previously the New Zealand Longitudinal Study of Ageing (NZ LSA). All ethnic groups. Postal survey then face-to-face interviews. Cognitive testing included (Kiwi ACE-R, which is modified for ANZ).
Life and Living in Advanced Age (LiLACS) [603]	Auckland University	Bay of Plenty & Lakes DHBs	2010-	80-90yrs in 2010	937	First longitudinal study of indigenous population ages >80 years (Maori and non-Maori participants). Cognitive testing included (Kiwi ACE-R).
Growing Up In NZ (GUiNZ) [601]	Auckland University	Auckland & Waikato	2009-2010	6-7yrs in 2016	6,846	Birth/longitudinal cohort. All ethnic groups included. Pre-natal recruitment included. Children, mums & partners included if child born between 2009-2010.
PREDICT-CVD [604]	Auckland University	Auckland, North Island	2002-2009	35-74yrs in 2002-2009	272,682	Research cohort established from CVD assessment tool in primary care 'PREDICT' integrated with electronic health records. Auckland initially, now other areas but mainly North Island.
The DUNEDIN study [605]	Otago University	Dunedin, Otago, South Island	1972-	43-44yrs in 2016	1,037	All ethnic groups included. Face-to-face assessments (biomedical and psychological).

Table 4.1: 2016 Longitudinal cohort studies in ANZ

Discussions with researchers, and meeting with a contact in the Ministry of Health (MoH) in Wellington, realised the potential for using the Integrated Data Infrastructure (IDI) [606], a whole-of-population, routinely collected data source for public-good research, to explore the geographical variations in dementia in ANZ. I now discuss this dataset (IDI) in more detail.

4.6.2 Integrated Data Infrastructure (IDI)

4.6.2.1 General overview

ANZ has, like Scotland, a unique identifier called the National Health Indicator (NHI). This has provided opportunities to link data. Therefore, ANZ has de-identified, linked health and administrative data for individuals available at a population level and available within a Trusted Secure Research Environment (TRE) called the ‘data lab.’ Together, all the routinely collected data is known as the Integrated Data Infrastructure (IDI), which was established in 2011. Statistics New Zealand are the data stewards for these data [494, 600]. The IDI is: *“a large research database containing microdata about people and households. Data is from a range of government agencies, Statistics NZ surveys including the 2013 Census, and non-government organisations [606]. The IDI holds over 166 billion facts, taking up to 1.22 terabytes of space – and is continually growing. Researchers use the IDI to answer complex questions to improve outcomes for New Zealanders”* [494, 600, 607]. Figure 4.1 highlights the many databases and sources of information for a person within the IDI.

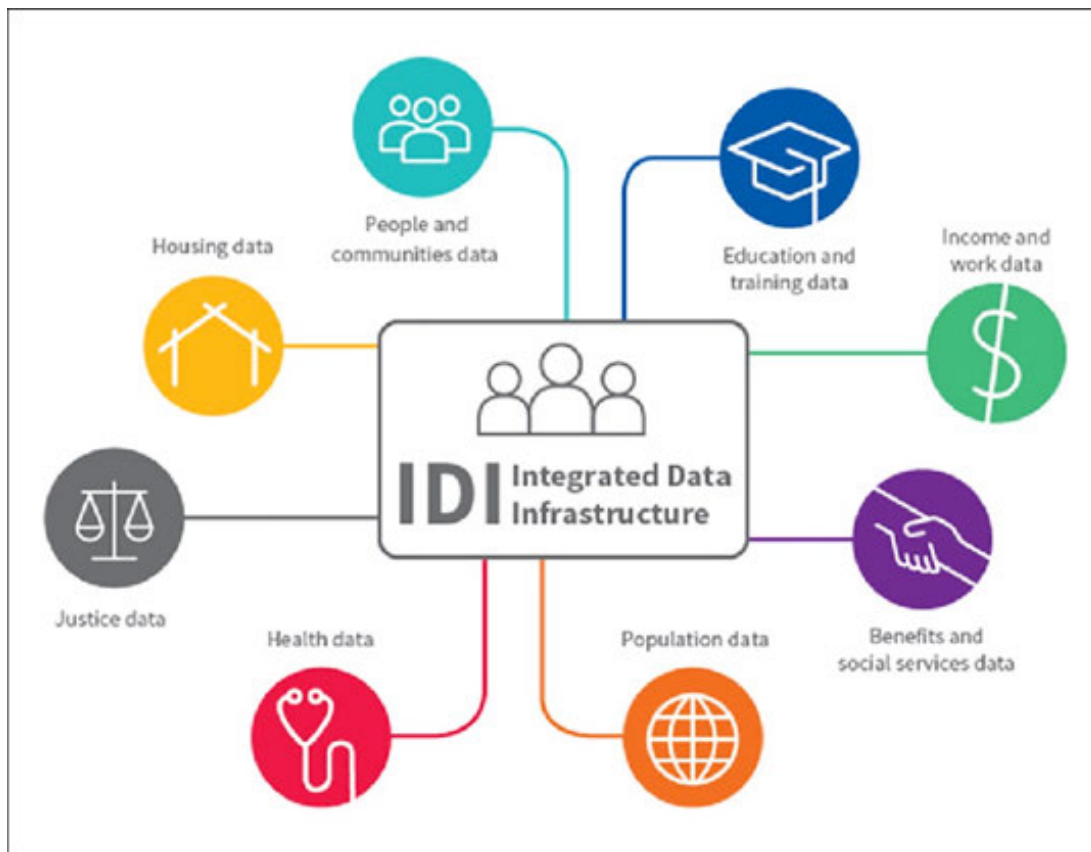


Figure 4.1: Datasets within the IDI for researchers [608, 609]

The IDI contains person-centred microdata from a range of government agencies, Statistics NZ surveys including the 2013 Census, and non-government organisations. This infographic image of the IDI datasets is from 2024 but represents the datasets available as of my research in 2016.

Source: Data in the IDI. Author: Statistics New Zealand, 2024. Used with permission under Crown Copyright and Creative Commons Attribution 3.0 New Zealand licence.

The key component of the IDI is the ‘Spine’, a unique identifier (called the `snz_uid` variable in the IDI, Stats NZ Unique identifier) which every person who has interacted with at least one government agency (health, education, social development, justice) in a given calendar year has been allocated by Statistics New Zealand. This allows different datasets to be linked to the ‘Spine.’ Researchers approved to use the IDI environment and particular datasets within the IDI may link data tables from different sources (e.g., health and education) they have ap-

proved access to, enabling research questions approved by data providers, Statistics New Zealand and ultimately the Government Statistician to be investigated. Figure 4.2 provides a schematic representation of using the Spine to link to four different datasets (referred to hear as ‘Nodes’).

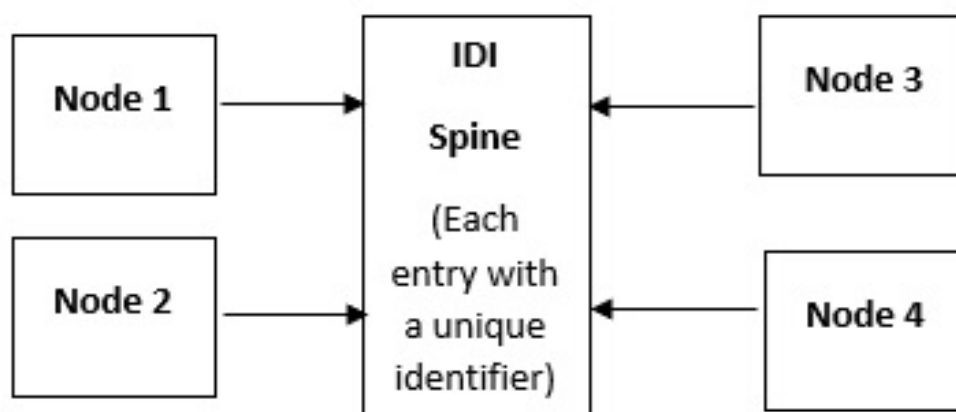


Figure 4.2: The IDI ‘spine’ and different nodes adapted from The IDI prototype spine’s creation and coverage by A. Black (2016) [600]

Once researchers have approval for their project from Statistics New Zealand (and also appropriate ethics committees) the research is given access to the IDI within the TRE. I discuss this approval process with a timeline in [Timeline of developing the ANZ IDI dementia project](#). Approval to the IDI provides the researcher access to a Microsoft Structured Query Language (SQL) Server database and selected tables relevant to the research question(s) within the TRE. The SQL is used by researchers to interrogate the IDI datasets and create a cohort with `snz_uid` unique identifiers as the primary key.

4.6.2.2 Preparing to use the IDI: access, permissions, ethics and governance

As discussed in [Chapter 3: Using big data for dementia research](#) there are variations in different countries with permissions and access to routinely collected data. ANZ uses two frameworks for this which are 1) the Five-Safes Framework, which are used for routinely collected data as discussed in [Chapter 3](#), and also 2) ‘Nga Tikanga Paihere’ [[610](#), [611](#)]. This second framework is specific to ANZ and involves five principles which are related to ten tikanga (Te Ao Māori, Māori world concepts) aimed to establish goals, boundaries and principles to guide and inform data practice [[610](#), [611](#)]. Full details on this are shown in [Figure 4.3](#).



Figure 4.3: Outline of the Ngā Tikanga Paihere framework: five central principles linked to two tikanga [610, 611]

Source: Ngā Tikanga Paihere Guidelines. Author: Statistics New Zealand, 2020. Used with permission under Crown Copyright and Creative Commons Attribution 3.0 New Zealand licence.

In summary to access the linked data within the Integrated Data Infrastructure in Aotearoa New Zealand I applied to Statistics New Zealand to access the Trusted Research Environment (TRE) at the University of Auckland. Full details of the process and permissions are outlined below and in Table 4.3 ([Timeline of developing the ANZ IDI dementia project](#)).

Access and permissions

At the time of my research (2016-7), all researchers had to be in ANZ to undertake their research. This required a further trip to ANZ in 2017. Researchers must also undertake their research in a TRE called the ‘data lab’ using only those computers. These were available at Statistics New Zealand’s regional offices and ‘approved’ University sites. The Data lab is completely locked down to external entities and therefore analysing data virtually using a VPN or other virtual system was impossible. Moreover, the law in ANZ, at that time, prevented researchers analysing data outside of the country.

The access and permissions process involved completing a Statistics New Zealand ‘pre-application’ interview and subsequent completion of an application form with review from Statistics New Zealand at several levels before full ‘sign off.’ Additional training for researchers not familiar with the IDI was required for meeting the ‘safe researcher’ principle of the Five-Safes Framework. I underwent this virtually with a Statistics New Zealand representative.

Ethics

Whilst informed consent was not obtained per rule 11(2)(c)(iii) of the NZ-Health Information Privacy Code [612]. This allows for anonymized health data research, an application to the Health and Disabilities Ethics Committee (HDEC) New Zealand [613] was required for a ‘Scope of review’ to assess if it did not require full ethics approval. Subsequently, I received confirmation from HDEC that my project did not require submission for a full HDEC review. I received an ‘Out of scope’ letter issued by HDEC to confirm this.

During discussion at the University of Auckland, whilst in ANZ, it became clear that it was necessary to also have approval from the University ethics committee - University of Auckland Human Ethics Committee (UAHPEC). There had initially been some ambiguity around this which I had sought to clarify. Once it became

clear, I did not proceed with the research. I did not start the project until this was processed by the University of Auckland's ethics committee whilst I was in ANZ.

Data Sovereignty

One important aspect of Data Sovereignty that I became aware of during my visits to ANZ, and have since learnt more about, is recognising the importance of Indigenous Data Sovereignty and Māori Data Sovereignty. There are important differences with a key aspect being that: *"Māori data should be subject to Māori governance"* [614, 615]. It refers to Māori and Iwi (Māori tribes/collection of people) aspirations and wishes for the use of their data [614, 615].

As outlined in the Te Mana Raraunga (Māori Data Sovereignty Network) there are important distinctions between the three references to data sovereignty [614, 615]. Table 4.2 incorporates the definitions from the Te Mana Raraunga [614, 615]. In 2015, the Te Mana Raraunga developed from a meeting of Māori researchers and practitioners at a workshop by the Academy of the Social Sciences in Australia on Data Sovereignty for Indigenous Peoples.

This workshop considered the: *"the implications of the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP) for the collection, ownership and application of data pertaining to indigenous peoples and what these might mean for indigenous peoples' sovereignty"* [614, 615]. Since my research in ANZ in 2017, there has rightly been a substantial increase in awareness of the essential need for Māori Data Sovereignty and the social licence for data use in research, including research using routinely collected health data.

There is no universal definition of the term 'social licence' [616, 617, 618]. The concept relates to society's ability to give or withhold approval for an organisation/business to operate based on society's trust in that organisation to behave in a legitimate way [616, 617, 618]. Statistics NZ is the data custodian for the

IDI and needs a social licence to operate effectively. In 2018, trust in Statistics NZ was measured [619]. This showed that: 34% of New Zealanders had a positive or very positive attitude towards Stats NZ, 6% had a negative or very negative attitude, and 61% had no opinion towards the agency [619]. Further work is progress as it has been claimed that the social licence could be classed as ‘tenuous’. Social licence of the IDI, and ensuring Māori Data Sovereignty within the IDI are important developments in ANZ since my published manuscript in this Chapter.

Data Sovereignty type	Definition
Data Sovereignty	Data are subject to the laws of the nation it is stored in.
Indigenous Data Sovereignty	Data are subject to the laws of the nation from which it is collected
Māori Data Sovereignty	Māori data area subject to Māori governance. <i>"Māori data sovereignty supports tribal sovereignty and the realisation of Māori and Iwi aspirations."</i>

Table 4.2: Te Mana Raraunga (Māori Data Sovereignty Network) - Table created from text extracted from the Te Mana Raraunga documents [614, 615]

Timeline of setting up my research in ANZ within the IDI

The timeline of setting up the project is highlighted in Table 4.3. This is helpful when later contrasting the longer timeline for setting up my research project in Chapter 5: Methodology for the SMS1947 dementia data linkage study in Scotland in Table 5.6.

Timeline	Tasks
May 2016	Initial scoping visit to New Zealand with MRC PhD student Travel grant to meet researchers and clinicians.
8th June 2016	Initial video meeting with Statistics New Zealand to establish if my project plan was feasible – ‘pre-application meeting’.
9th Nov 2016	Pre-application interview video meeting with revised plan.
Nov 2016	Statistics New Zealand application for review.
Jan 2017	Submission of Statistics New Zealand application.
13th Jan 2017	Submission of ‘Scope of review’ application form to Health and Disabilities Ethics Committee (HDEC) New Zealand.
16th Jan 2017	Confirmation from HDEC that study does not require submission for full HDEC review. An ‘Out of scope’ letter issued by HDEC.
15th Feb 2017	Confidentiality training video meeting to allow access to the Integrated Data Infrastructure.
15th Feb 2017	Initial approval from Statistics New Zealand (approved on subject matter, legal and methodology teams). Final approval awaited.
February 2017	Signing contract with Statistics New Zealand to abide by researcher requirements.
2nd Mar 2017	Final approval of project within IDI by Government statistician at Statistics New Zealand.
6th March 2017	Return to New Zealand to undertake the research in IDI in University of Auckland’s data lab.
13th Mar 2017	Research associate status with University of Auckland and login access to the IDI in the data lab in University of Auckland.
15th Mar 2017	Suspended any access to data lab whilst discussions with University of Auckland ethics committee to clarify whether additional need for local University of Auckland ethics committee approval first
5th Apr 2017	Application submission to University of Auckland Human Ethics Committee (UAHPEC) for ethics approval.
6th Apr 2017	Approval from UAHPEC of ethics application. Ethics approval for 3 years.

239
Table 4.3: Timeline of developing the ANZ IDI dementia project

4.6.2.3 Datasets available within the IDI

It was only possible to discuss the methodology of my research using the IDI to investigate dementia prevalence and geographical variation briefly in the published article (Section 4 4.3), so this section provides additional information relevant to the research and using datasets within the IDI to construct the whole-of-population dementia prevalence.

I had hoped to add memory clinic data, available as a dataset outside the IDI, into the IDI. Although this was theoretically possible, it was not a priority for Statistics New Zealand, and was not possible at the time of my analyses and visits to ANZ. Therefore, I used the datasets on hospital admission, drug prescribing, and deaths, which were available within the IDI, and were the most appropriate for dementia ascertainment.

Dataset	Date	Source	Description
Pharmaceutical	2005-	MoH	ANZ pharmaceutical funding agency (PHARMAC). Subset of the Pharmaceutical Claims Collection. It contains information about subsidised dispensing of drugs.
Publicly funded hospital discharges	1988-	MoH	Subset of fields from the National Minimum Dataset (NMDS). Includes discharge and event data (event/diagnosis/procedure) about publicly funded hospital events and demographic data reported for the population cohort.
Mortality	1988-	MoH	Data of underlying cause of death for all deaths registered in New Zealand, using the World Health Organization Rules and Guidelines for Mortality Coding.

Table 4.4: Electronic datasets used within IDI to ascertain dementia for 2012-2015 [609]

4.6.2.4 Pharmaceutical dataset

Based on a cost-benefit analysis, ANZ's pharmaceutical funding agency (PHARMAC) only publicly funds two dementia medications (Donepezil and Rivastigmine) at the time of my research. Rivastigmine has only been available since 2012 onwards in ANZ, prior to that only Donepezil was publicly funded. The other two available as anti-dementia drugs are Galantamine and Memantine. During the years my research investigated, these were only possible if self-funding the treatment, which ranged from NZ\$150-300 a month. Dementia drugs now available in ANZ include Donepezil, Rivastigmine, Galantamine, and Memantine (privately self-funded) [620, 621, 622]. On discussion with clinicians in ANZ who worked in memory clinics during my research time, they confirmed that Galantamine and Memantine were very rarely self-funded.

Given this information, for the purpose of ascertaining dementia, I only used Donepezil and Rivastigmine. Rivastigmine was only available in ANZ from 2014 onwards, so not available for ascertaining dementia in 2012-2013 data extraction. At the time of my research it was not clear the percentage of people with dementia in ANZ who were prescribed drugs, although there was evidence of variation in prescription of anti-dementia drugs [623]. Further research since my publication has shown on-going variation [620].

A list of the drugs and their PHARMAC ID code and drug class is outlined in the supplementary material for my published paper in Section 4.3.1 ([Supplementary manuscript information – not peer reviewed](#)). This is also reproduced in Table 4.5 ([Anti-dementia drugs used for dementia ascertainment in ANZ, 2016](#)) for clarity in this section.

Drug	PHARMAC ID	Drug class
Donepezil hydrochloride	392326, 392325	Acetylcholinesterase inhibitor
Rivastigmine	403725, 403726	Acetylcholinesterase inhibitor

Table 4.5: Anti-dementia drugs used for dementia ascertainment in ANZ, 2016

4.6.2.5 Publicly funded hospital discharges and mortality data

Both these datasets are from the Ministry of Health. In my published paper in Section 4.3 I outline these two datasets.

The classification systems used in New Zealand hospitals are the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM [624]. Additionally, the Australian Classification of Health Interventions (ACHI) and Australian Coding Standards (ACS) are included [624]. Collectively the full classification is referred to as ICD-10-AM/ACHI/ACS and the Australian Refined Diagnosis Related Groups (AR-DRG) [624]. I only used ICD-10-AM codes for extracting dementia for consistency. I discussed using the ICD-10 AM (Australasian modification of the ICD version 10) dementia codes for ascertaining dementia in the published paper [3]. The codes were available in the online supplementary appendix data (in Section 4.3.1 for the published paper). For ease, I have also listed them in Table 4.6.

ICD 10	Description
A810	Sporadic Creutzfeldt Jakob disease
F051	Delirium superimposed on dementia
F00	Dementia in Alzheimer disease
F000	Dementia in Alzheimer disease with early onset
F001	Dementia in Alzheimer disease with late onset
F002	Dementia in Alzheimer disease, atypical or mixed type
F009	Dementia in Alzheimer disease, unspecified
F021	Dementia in Creutzfeldt-Jakob disease
F024	Dementia in human immunodeficiency virus (HIV) disease
F022	Dementia in Huntington disease
F02	Dementia in other diseases classified elsewhere
F028	Dementia in other specified diseases classified elsewhere
F023	Dementia in Parkinson disease
F020	Dementia in Pick disease
F013	Mixed cortical and sub-cortical vascular dementia
F011	Multi-infarct dementia
G308	Other Alzheimer disease
F018	Other Vascular dementia
F012	Sub-cortical vascular dementia
F03	Unspecified dementia
F03X	Unspecified dementia
F01	Vascular dementia
F010	Vascular dementia of acute onset
F019	Vascular dementia, unspecified
F106	Mental and behavioural disorders due to use of alcohol - amnesic syndrome
G30	Alzheimer disease
G300	Alzheimer disease with early onset
G301	Alzheimer disease with late onset
G308	Other Alzheimer dementia
G310	Circumscribed brain atrophy
G311	Senile degeneration of brain
G318	Other specified degenerative diseases of nervous system
I673	Binswanger's disease

Table 4.6: ICD 10-AM diagnostic codes used for dementia ascertainment in ANZ

4.6.2.6 Data processing in the IDI

At the time of my research 2017-2019 in the IDI, research had been accepted for publication following peer-review by two of my co-authors on methods in how to construct a population cohort within the IDI which had been found to be representative of ANZ's population [625]. For the purpose of my research we used these methods, with a focus specifically on dementia. This involved creating two groups: 1) a 'denominator' (those at risk of dementia or the *study*) population of all in ANZ ≥ 60 years (alive or died during the year) annually for years 2012-2015, and 2) a 'numerator' group which was those with dementia.

To undertake the first step of creating the 'denominator' population within the IDI which represented the population of ANZ between 2012-2015, we selected all individuals ≥ 60 years who had interacted with any of the datasets listed in Table 4.7 at any point annually for 2012-2015 within the IDI. This included interaction with any of the health datasets: primary health Organisation (General Practice registration), publicly funded hospital discharges (part of the National Minimum Datasets (NMDS), subsidised dispensing in the Ministry of Health (MOH) and Pharmaceutical Management Agency (PHARMAC), and mortality data from the Ministry of Health. The administrative datasets used for the creation of the denominator (*study*) population were administrative datasets: Inland Revenue (pensions), and Accident Compensation Corporation (ACC). This created an annual *cohort* for each year 2012-2015 representing those in ANZ ≥ 60 years.

Traditionally, in research, ≥ 65 years is when dementia is deemed 'late onset.' However, on discussions with clinicians in ANZ we changed our age group to anyone ≥ 60 years to prevent exclusion as it was yet unclear what was appropriate for a multi-ethnic country with an indigenous population. Subsequently, research was published suggesting that dementia may present at a younger age in the Māori population [626].

Dataset	Type	Description
Inland Revenue [609]	Administrative	This dataset incorporates pensions data. Anyone retired and receiving a pension in ANZ would be listed in this dataset. This was a useful dataset to capture the population who were retired and receiving pensions in ANZ annually for the denominator (<i>study</i>) population for the study.
Accident Compensation Corporation (ACC) [627, 609]	Administrative	Everyone in ANZ is covered by ACC's no-fault scheme if they are injured or have an accident. The cover helps to pay towards the recovery (i.e. physiotherapy) following an accident. It is a helpful dataset for capturing individuals within the population, so was used for creating the annual denominator population. On discussion with clinicians specialising in dementia, and a review of the data in the IDI, it did not have many codes for dementia that were not found in other datasets. Given the context of dementia, those with injuries would likely feature in other datasets used for dementia ascertainment in the study.
Primary Health Organisation [609]	Health	Data from 2003 in the IDI. Contains information for those enrolled in a primary health organisation (general practitioner), including their demographics, when they enrolled, when they were last seen and the type of primary care practice type.
National Minimum Datasets (NMDS) [609, 628, 629]	Health	This comprises secondary care data including information on hospital admissions/discharges. It was implemented in 1993 with public hospitals in ANZ submitting data from 1993 (although it was backloaded with data to 1988), and private hospitals from 1997. The NMDS contains clinical information including patient's diagnoses, procedures, length of stay, and patient demographic information. The Publicly Funded Hospital Discharges is one component of this dataset. The diagnoses are coded to ICD codes (the Australian Modification version of the ICD).
Pharmaceutical data (also referred to as PHARMAC) [609]	Health	The Pharmaceutical data includes a subset of fields from the Pharmaceutical Claims Collection, and particularly subsidised dispensing of drugs. It includes demographic information about healthcare users.
Mortality [609]	Health	This comprises of death certificate information. Diseases contributing to death are listed as ICD codes (Australian Modification Version). It is a Ministry of Health dataset. There is a temporal lag in release of this death certificate, to all any deaths that have required the coroner's input to be included. This dataset was used to capture anyone who had died during the year.

Table 4.7: Electronic datasets used within IDI to create annual denominator (*study*) populations for 2012-2015 (using the IDI October 2016 version)

The next step was to create the ‘numerator’ cohort i.e., those with dementia. Datasets within the IDI that would be appropriate for ascertaining dementia were selected, following stakeholder engagement about the diagnosis of dementia with a number of data scientists and clinicians. I also sought expert guidance from Dr Phil Wood, a geriatrician in ANZ involved in dementia clinics and Chief Advisor for Health Ageing at the Ministry of Health, who subsequently became a co-author on the manuscript.

As highlighted in my published paper on this research (Section 4.3), the datasets in the IDI at the time of my research that were felt to be most beneficial for ascertaining dementia were the pharmacy dataset (PHARMAC), mortality and public hospitals discharges datasets - a subset of the National Minimum Dataset (NMDS) dataset. These are outlined in the published manuscript in this Chapter, but I highlight these again in Table 4.4 for ease. One external dataset (outside the IDI safe haven) that was in the process of being included into the IDI was interRAI [630]. This would have provided a valuable resource for anyone with dementia. However, this process was going to take longer than my research time frame in ANZ allowed. This has now been linked, since my publication, to the IDI and I discuss this in the section on research since my publication as ANZ researchers have expanded using the IDI for dementia prevalence with this additional dataset since my seminal paper.

Lastly, I wanted to triangulate the dementia diagnoses within the IDI with a dementia specific dataset. Following discussion with Dr Wood, I asked Statistic NZ if we could link an external dataset (a large memory clinic data from an Auckland memory clinic at North Shore public hospital) to the IDI. Whilst it was theoretically possible (with appropriate permissions and ethics in place) the time-frame would again be outside my time-frame for my research period unfortunately. There are other datasets related to health in the IDI, but when I reviewed them, they were not appropriate for dementia ascertainment. Unfortunately, at the

time there was no primary care data available in the IDI with health information for dementia ascertainment. These are some of the limitations I highlight in my published paper (Section 4.3) and in Section 4.10 ([Additional discussion and future work](#)).

Data output from the IDI

Alongside being the data stewards [631] for the IDI, Statistics NZ also check all research results prior to release ('output') to ensure individuals cannot be identified. They have systems and processes in place to further ensure this and researchers must use these prior to requesting output of their data from the secure research facility or Trusted Research Environment (TRE). These include suppressing low numbers (counts <6) which means they are outputted only as 'S' (Suppressed) and not the actual number. The additional process that Statistics NZ ensures all researchers undertake on the data prior to output is called 'random rounding to base of 3' (RR3) [632]. What this means is that the actual numbers outputted are very slightly different to the actual number the researcher found in the IDI analysis. It adds an additional layer of ensuring no one can be identified.

For example, a count of 5 is displayed as a 3 or 6, and a count of 1 is displayed as a 0 or 3. Statistics NZ helpfully describes this process: *"The counts are randomly rounded to base three in a consistent manner. This is to disguise small counts, but all cells in the table are randomly rounded. The effect is to make the output more confidential, by generally preventing individuals' data from being released"* [632]. I have included Statistics NZ information sheet on this in [Appendix 3](#). The benefit of this method is seen most when numbers are small. I later discuss in my research in Scotland ([Chapter 5: Methodology for the SMS1947 dementia data linkage study](#)) using routinely collected health data that numbers less than 6 are suppressed in output. The approach in Scotland is different, but consistent with the suppression rules in ANZ.

The effect that RR3 has on the final output is clarified by Statistics NZ: *"For*

small numbers, where there is the most risk that individuals could be identified, there are larger percentage changes compared with larger numbers. For example, a cell with a one changed to a three has been changed by 200 percent, but a cell with 1,001 changed to 1,002 has been changed by only 0.1 percent. When analysing data, small counts need to be treated with caution but for the larger values the percentage changes in these cells do not cause a problem” [632].

4.7 Geographical scales in ANZ

ANZ is divided into North and South Island and within that there are various geographical scales based on administrative data such as census data or local government levels, or health areas. Unfortunately, ANZ’s postcode geography is not as fine as the UK. A single postcode could cover large geographic areas across urban and rural parts of a region with very large populations. Essentially, the ANZ postcodes are more similar in size to Scottish postcode sectors and/or districts, and are not appropriate for small area data analyses.

Until 2018, the smallest geographical unit used for statistical output of census data in ANZ was the Meshblock (MB), which had an average population of 100 people, but ranged from zero to approximately 2,000 [633]. Meshblocks vary in spatial extent, from a small part of a city block to large areas of rural land. As Meshblocks are small and typically have small populations, information about sub-populations at this scale is sparse due to the Stats NZ confidentiality constraints.

Census Area Units (CAU), which are also referred to as Area Units, were designed as aggregations of Meshblocks and were akin to suburbs, with average populations of approximately 2,000 people and some areas having in excess of 5,000 people [634]. Importantly, the Ministry of Health used Census Area Units for the basis of their domicile units that were created to represent a patient’s usual residence.

Data zones are also used for analysis in ANZ, and were developed by the University of Auckland [635] as whole aggregations of 2013 Meshblocks for the base geography for the 2013 New Zealand Index of Multiple Deprivation (IMD) [636]. Similar to Scotland, Data Zones in ANZ have a population range of 500 to 1,000 [635, 636]. In 2013, there were an average of 712 people per Data Zone, and were designed to be large enough to enable the output of data from the IDI, while being small enough to notionally be neighbourhoods [636].

Depending on the nature of the disease, this may be identifiable if there are too few people with a disease in a small geographical area. Data Zones typically include eight census Meshblocks, and in urban areas may comprise a few streets. Again, this may be identifiable if the disease has a low population prevalence.

At the time of my analyses, publicly funded health services in ANZ were separated into District Health Boards (DHBs) [637]. There were twenty DHBs in ANZ, each with varying population sizes, ethnic compositions and health needs. IN ANZ, patients wishing to visit a General Practitioner (GP) must register with their practice at least once every 3 years. The selection of GP is not constrained to the location of their residence. Patients may choose to register with a clinic closer to their workplace than their home, or if they have relocated within a city such as Auckland, there is no requirement for patient to move to a more local practice. Table 4.8 summarises the geographical scales available at the time my research was carried out, 2016-2017.

Scale	Number	Available in IDI	Description
Meshblocks (MBs) [633]	46,637 in 2013 Census	Yes	This is the smallest geographic unit used by Statistics NZ for collecting and processing statistical data. It is defined by a geographic area and can vary in size from part of a city to larger rural area. All meshblocks form a network covering NZ. All meshblocks can be added together to build up larger geographic areas or to define electoral districts/regional councils or territorial authorities.
Census Area Units (CAUs) [634]	2,000 approx. in 2013 Census	Yes	Aggregations of Meshblocks. Similar to suburbs. The average populations are approximately 2,000 people. Some areas have in excess of 5,000 people. The Ministry of Health used Census Area Units for creating their domicile units to represent a patient's usual residence.
Data Zones (DZs) [636]	5,958 in 2013 Census	Yes	Small areas which include aggregates of approximately eight census meshblocks per data zone. The land mass was divided into 5,958 data zones based on 2013 census. Each data zone ranged from 501 to 999 people with an average population of 712, with the exception some data zones representing larger areas or smaller areas.
Postcodes [638]	1,856	No	Consists of four digits. The first two specify area, the third the type of delivery, and final digit for the specific lobby, RD number, or suburb.
Territorial authority (TA) [639]	67	Yes	A city or district council area. The second tier of local government below regional councils. The 67 TAs comprise 13 city councils, 53 district councils, and the Chatham Islands Territory.
District Health Boards (DHBs) [637]	20	Yes	Established in 2020. Initially 21 DHBs, but reduced to 20 in 2010.
Regional councils	16	Yes	Top tier of local government
North and South Island	2	Yes	By combining smaller areas above, can aggregate to North or South Island

Table 4.8: Geographical scales in ANZ 2015-2017

Figure 4.4 ([ANZ map with District Health Boards 2015](#)) includes two maps and highlights the layering of different geographical scales and their boundaries. On the left-hand side, is a map of Aotearoa New Zealand with District Health Board (DHB) boundaries from 2015 (clipped to the mainland and excluding water bodies). On the right-hand side is a *"zoomed in"* map of the Auckland Central Business District (CBD) which shows the spatial boundaries of the DHB in 2015, the CAU 2013, the DZ 2013 and MB 2013. Details of the creation of this, including the spatial boundaries of the different geographical scales, are in [Appendix 4](#).

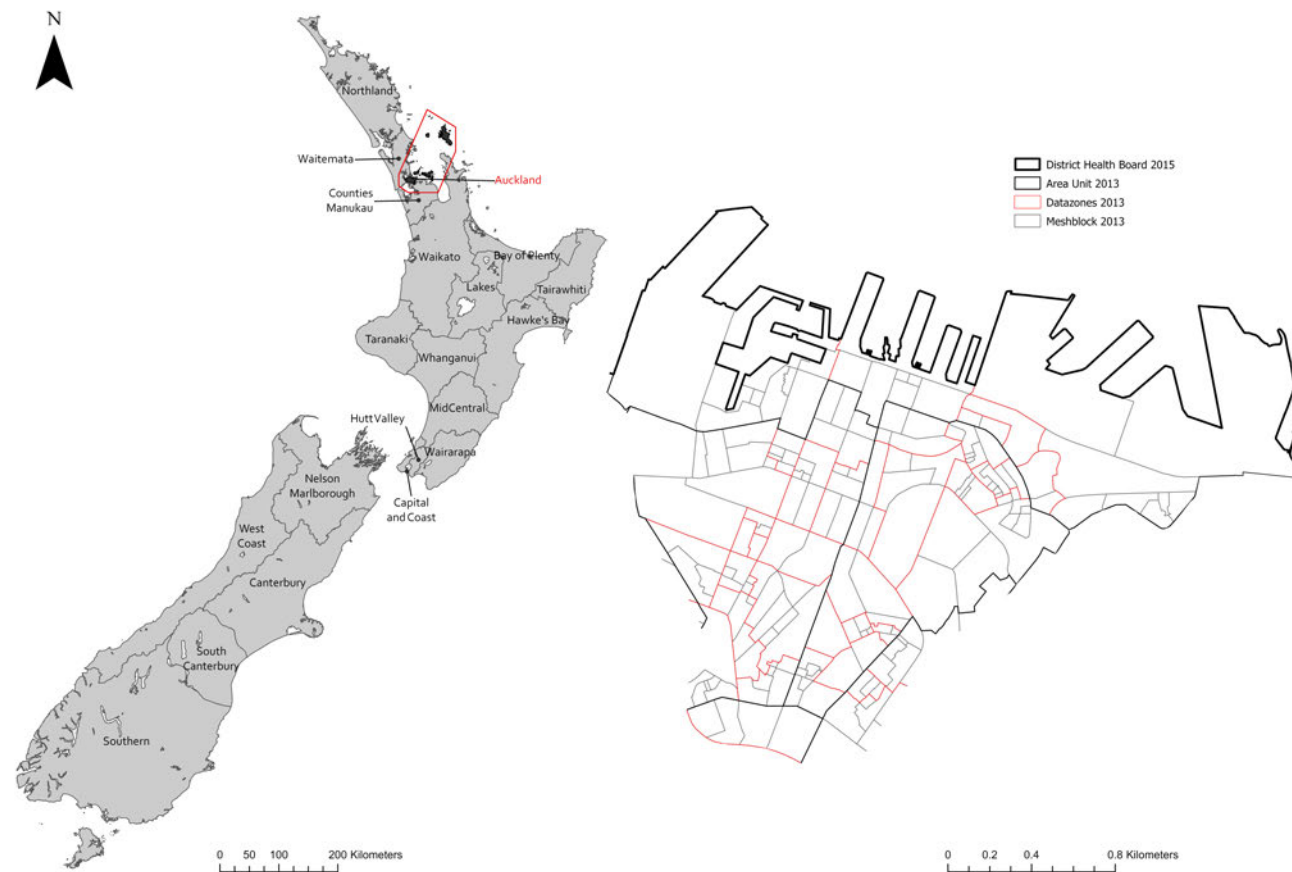


Figure 4.4: ANZ map with District Health Boards 2015 on the left, and a "zoomed-in" view of Auckland Central Business District with different layers of geographical scales on the right. Scale bar for each point of reference. Auckland DHB is coloured black.

Source: Jessie Colbert, University of Auckland, 2024. Used with permission.

From 2018, Statistics NZ introduced the Statistical Area 1 (SA1) to replace Meshblocks, with a tighter control on the population counts in each SA1 of between 100 and 200 people [640]. Statistical Area 2 (SA2) units, replaced CSUs and have similar population constraints to SA1s, with a target population of 2000 to 4000 people [641]. Meshblocks have been retained by Statistics NZ as the base geographical unit for building higher output geographies and can still be used by researchers in the IDI [642].

I initially analysed the data using different geographical scales, but ultimately exported data at the largest geographical scale: North and South Island. This was due to practical factors (i.e., only being able to do analyses while on site in ANZ, and a limited time period in ANZ), and factors out with my control (closure of the datalab due to Cyclone Cook in 2017).

Deprivation

In addition to the geographical scales mentioned above, there are deprivation scales in ANZ. This includes the New Zealand Index of Multiple Deprivation (NZ IMD or IMD13) which was created in 2013 in a similar structure to the Scottish Index of Multiple Deprivation (SIMD) [643]. The NZ IMD incorporates 28 indicators of deprivation categorised into seven Domains (Employment, Income, Crime, Health, Housing, Education, Geographic Access to Resources) [643]. There have been further developments since, and expansion in the number of indicators used to class an area. Each IMD version is named according to the year of update e.g., IMD18 for 2018. A new Index of Multiple Deprivation specifically considering deprivation for older people (The Older Person's Index of Multiple Deprivation) in ANZ [644] has been created since my publication and may benefit future work in understanding dementia prevalence and variations.

4.8 Impact of my research

4.8.1 How does this study add to existing literature?

Whilst there are excellent epidemiological studies of dementia prevalence in ANZ (including the first indigenous longitudinal cohort, LiLACs NZ) [603], none of these were nationally representative prior to my research. As highlighted earlier, during my first trip to ANZ I learned that previous estimates of dementia prevalence in ANZ were modelled on European and Australia cohorts and extrapolated to the population of ANZ. My research adds to the existing information by producing cautious estimates of dementia prevalence using objective health data for all ethnic groups between 2012-2015, whilst concurrently demonstrating the potential for the IDI as a resource for dementia research.

To my knowledge, this was the first national estimate of dementia prevalence using ANZ's own data for a whole-of-population cohort and not dependent on survey data. As highlighted in the peer reviewed publication [3], these estimates were lower than previously reported estimates for ANZ [592, 593], however, our research highlighted higher estimates for Māori. This builds on growing research highlighting that late onset dementia might present earlier (<60 years) in this population group [645, 646].

The findings from my research give important context building on past research with three key points to consider: 1) dementia prevalence in ANZ and how country-specific estimates are created representing the ethnic variation of ANZ; 2) cultural differences in seeking a diagnosis of dementia or not; and 3) growing awareness that late onset dementia may present earlier in different ethnicities within ANZ.

4.9 Strengths, limitations and challenges

One of the strengths of my research was that it was the first published paper to use ANZ data to estimate ANZ dementia counts in a set time period. There are caveats to this, as highlighted in the publication, that these estimates may be cautious estimates and subsequently further research has been undertaken including additional datasets I suggested in my publication, but were unavailable to be linked at the time of my research.

4.10 Additional discussion and future work

My research sets the foundation for additional research studies investigating dementia in ANZ at a population level. Another study expands my seminal research of dementia prevalence estimates using the IDI to include a published study protocol to estimate dementia prevalence, with the inclusion of the InterRAI data that has recently become available within the IDI. Furthermore, there are future plans for a detailed national dementia prevalence study in ANZ.

The evolution of the IDI for whole-of-population cohort analysis has become apparent, with my research and subsequent studies demonstrating the IDIs utility for dementia prevalence studies to analyse cardiovascular disease in the IDI [604, 625]. Since my publication, there is growing research in ANZ investigating dementia (including within different ethnic groups) [646, 647, 648, 649].

One study investigated different methods for diagnosing dementia in the Māori population of Aotearoa New Zealand by using the 10/66 dementia protocol (traditionally used in Low to Middle Income Countries as I discussed in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#)) [649].(cite) Another study investigated this 10/66 protocol in the Fijian-Indian population in ANZ [648]. Using a clinical diagnosis as the reference standard, the study concluded that the 10/66 diagnosis had a sensitivity of 92.3% (95%CI 70.3-

99.5), specificity of 93.8% (CI 75.3-99.6) and positive predictive value of 92.4% (95%CI 70.3-99.5) [650].

A further study has investigated the provision of anti-dementia drugs and whether this varies amongst different ethnicities (something I highlighted in my discussion in my publication). That research found that fewer acetylcholinesterase inhibitors were dispensed in South Auckland amongst Māori and Pacific people [646]. There is also an updated Dementia Financial report [568] in 2021 since my research with estimates for future dementia predictions, and also increased input from ANZ dementia researchers in driving forward the need for more country-specific research.

Recognising the role of national routinely collected data for dementia research in ANZ is a positive step. On my first visit to ANZ, my research idea was proposed to a large national funding body to obtain a grant. This was unsuccessful, but it was clear that there was caution, at that time, in using routinely collected data within the IDI to answer dementia prevalence questions rather than the more traditional epidemiological research studies. I believe my publication, although a brief report, has been helpful in highlighting the IDI as a useful tool and a paper to reference for researchers in ANZ to build on developing a more detailed understanding of dementia prevalence there. Since publication, my code and methods for ascertaining dementia within the IDI have been used subsequently in four published papers investigating dementia [651, 652, 357, 653].

A substantial limitation of my project was the geographical scale that results could be outputted. Since publication of my research, there have been two important changes which need consideration for future geographical analysis of dementia prevalence in ANZ. First, the 2020 Labour-led government abolished the 20 District Health Boards (DHBs) and replaced them with a different, single agency called Te Whatu Ora (Health New Zealand) and a separate entity called Te Aka Whai Ora (Māori Health Authority) for health services for the Māori

community as of 2022.

These changes are being repealed following the election of a new coalition government in late 2023. In relation to this chapter, and the whole thesis, the changing scales of analyses highlights the need to determine what ‘place’ means for analysis in both time and space. In the context of this thesis, for example, does place refer to a geographical scale that is consistent over time (but with modifications of the services provided in that area that might influence health) or is it a changing unit over time that becomes difficult to compare and contrast over time.

Considering the factors that make up ‘place’ over time and over our lifetime and how that might influence our health well-being and risk of disease, there has been a publication of the Older Person’s Index of Multiple Deprivation in ANZ [644, 654], based on the NZ IMD, since my research. The aim is to measure the deprivation circumstance of older populations across ANZ, recognising that the NZ IMD might not be representative for older people living there. This highlights a growing awareness (not solely in ANZ but internationally) that location and how that influences our health, well-being and disease is important to understand at a country-specific level.

Finally, a report in 2023 building on the 2021-2024 Health Strategy in ANZ to understand ‘Older people experiencing vulnerability and multiple disadvantage in New Zealand’ [655] highlighted that for physical and mental health conditions there was an 8% points difference in health vulnerability between older people in urban areas compared to rural areas in ANZ. The report defined health vulnerability as two or more physical or mental health conditions. This report showed that 26% of older people in urban areas experienced health vulnerability compared to 18% in rural areas [655].

The authors stressed that this could indicate a gap in services for older people in rural areas rather than improved health in those areas [655]. Whilst the pre-

liminary analysis did include dementia and Alzheimer’s disease, the report did not while other mental health conditions were considered in the report. Figure 4.5 from the report illustrates regional variability in health vulnerability amongst older people in ANZ.

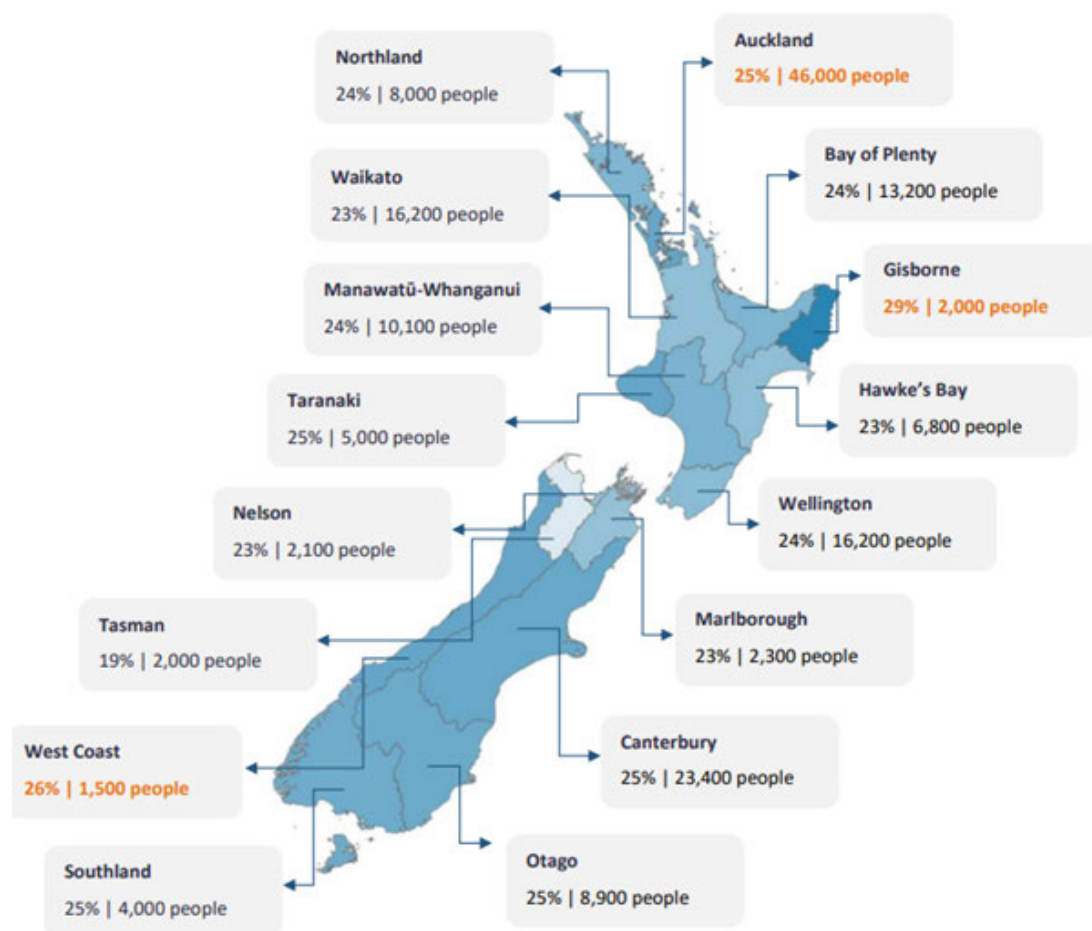


Figure 4.5: Regional variability of health vulnerability across ANZ for physical and mental health conditions (excluding dementia and Alzheimer’s disease) [655]

Source: ‘Older people experiencing vulnerability and multiple disadvantage in New Zealand: A report on the needs of older people (65+) in health, housing, finance, social connection, and access’ report [655] Authors: Sarah Ann Pinto, Penny Mok, and Verity Warn. Social Wellbeing Agency (2023). Used with permission under Crown Copyright and Creative Commons Attribution 4.0 International licence, creativecommons.org/licenses/by/4.0.

This is important context in building a country-specific level in understanding the needs and health of the older population living there. From a dementia perspective, given that age is the greatest risk factor for dementia, understanding the geographical spread of health, well-being, deprivation, alongside access to health services for those most at risk of dementia in ANZ, is important holistically to further dementia research. Adjusting for variables such as deprivation (NZ IMD), and known disparities geographically in health vulnerability are important additions to my research that I was unable to undertake at the time.

4.11 Chapter conclusions

A detailed national epidemiological study of dementia in ANZ is urgently needed to inform health policy and planning. A recent publication has expanded my work in the IDI to include the InterRAI dataset that I was not able to access at the time of my research. The complex interplay of who seeks a diagnosis, how this may vary between different ethnic groups, and whether different age points (specific to ANZ and its multi-ethnic population) are needed for late onset dementia, given research demonstrates possible earlier onset of dementia in Māori populations [645, 626].

As discussed in this chapter I used ≥ 60 years as my cut off for my population within the IDI to incorporate prior (not yet published at the time) emerging research that dementia might present earlier in Māori population [645, 626]. The traditional cut off for late onset dementia in research is ≥ 65 years. However, further work is needed for establishing the ‘best’ cut off for the denominator population in ANZ for late onset dementia research. For example, ANZ’s CVD guidelines [656, 657, 658, 659, 604] explicitly include Māori and Pacific peoples 10 years earlier than Pākehā (White New Zealander) in their eligibility for screening.

There is research into understanding the different geographical variations of diabetes and other cardiovascular diseases in ANZ [660], and this is also needed for

dementia, including a deeper understanding of all-cause dementia, Alzheimer's Dementia and Vascular Dementia within different ethnic populations so that appropriate healthcare services and potentially modifiable risk factors can be tailored to both population and individual level needs. Kaupapa Māori research methods are required to appropriately progress this research with a strengths-based lens for Māori. This is vital, given previous deficit-based research has typically created greater health inequalities and divisions.

As mentioned, I had intended to incorporate data from a large memory clinic located in Auckland City's North Shore public hospital. These data would have allowed triangulation of results and provided information on sensitivity and specificity. Patients attending the memory clinic with confirmed dementia could then be followed up to ascertain whether a) they were featuring in the pharmaceutical database with prescription of an anti-dementia drug, or b) whether their discharge letter had dementia recorded for patients who were subsequently admitted to hospital, or c) if their death certificate recorded dementia, when a patient had died.

While constraints, out with my control, meant that the linkage of this external dataset to the IDI data could not proceed within the time-frame of my research, it is still an important area of future research that could inform dementia service policy and planning and give country-specific information of the validity of dementia coding in routinely collected datasets within the IDI.

International evidence suggests that the likelihood of a patient having dementia recorded on their death certificate can vary, but that if it is recorded then the positive predictor value of the person having dementia is high. This is why using death certificates as a means of dementia ascertainment in routinely collected healthcare data research is an accepted method. However, this information is not, to my knowledge, known yet for ANZ, and particularly how many death certificates do not record dementia when someone has a diagnosis of dementia.

Including the memory clinic data linked to the IDI could also form a de-identified case cohort linked to routinely collected healthcare data providing valuable additional detailed information which is often not available in many routinely collected healthcare datasets.

Finally, the new development of the Dementia Mate Wareware Initiative (a national dementia strategy) [[661](#), [590](#)] will hopefully be a powerful tool to aid future research work as well as important clinical care and services for people living with dementia in ANZ.

Chapter 5:

Methodology for the SMS1947 dementia data linkage study

5.1 Chapter introduction

In the thesis so far, I have:

1. Systematically reviewed the literature investigating rural versus urban living and risk of dementia;
2. Described routinely collected health data and discussed using it to answer my thesis question;
3. Presented my study investigating dementia between North and South Island in Aotearoa New Zealand (ANZ) using population level routinely collected data.

In this Chapter, I build on this earlier work by considering how to design and undertake a research project combining a historical survey from 1947, the Scottish Mental Survey 1947 (SMS1947), with routinely collected health data to investigate the association of rural or urban living and a diagnosis of dementia. I will discuss the methodology in this chapter and [Chapter 6: Results of the SMS1947 dementia data linkage study](#) reports the results with discussion.

As discussed in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#), the analysis I undertook in Aotearoa New Zealand was a ‘snap shot’ in a specific time period (2012-2015). It was not possible to undertake more longitudinal analysis for that research. One opportunity for capturing the benefits of the *breadth* of routinely collected health data with the *depth* of information available within a detailed study (e.g. research study,

longitudinal cohort, survey etc.) for healthcare research is to link the two through data linkage. A research study can be *nested* within routinely collected data with appropriate permissions. In [Chapter 3: Using big data for dementia research](#) I discussed the growing examples of linking research studies or population studies to routinely collected health data.

The methodology for the research in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) is discussed in this Chapter, including the stages involved in creating a research study combining a historical survey (SMS1947) with routinely collected health data. Within these stages, I give an overview of key information in Scotland relevant to this study design and an overview of the SMS1947. I then discuss the changes in methodology and study design that were required during the planning of this research. I highlight the planning, processes, and preparation to access the linked data to answer my research question. Following that I outline the methods used for the final study design.

5.2 Research aims and objectives

Given the developing interest in variation of dementia geographically as outlined in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#) of my thesis, I sought to investigate this further in Scotland. My initial research question was: *"Is there a difference dementia rates between different geographical areas of Scotland in the SMS 1947 linked to routinely collected health data?"* Following changes (which I will outline in this Chapter) my research question was adjusted to focus specifically on rural versus urban living: *"Is there a difference between rural and urban living and dementia rates in Scotland within the historical SMS1947 CHI-linked to routinely collected health data?"*

The aim of my study was to establish whether living in a rural or urban area had an impact on dementia outcome across Scotland by using a historical population-level dataset from 1947 (with information on childhood intelligence) linked to

routinely collected health data to investigate my research question.

The objectives of my research were fourfold:

1. To create a study population from the database population (the historical SMS1947 CHI-linked to routinely collected health data) within Scotland's National Safe Haven;
2. To ascertain who had dementia within this study population and the database population;
3. To ascertain a rural or urban living variable (using the Urban-Rural 2020 Classification system) within the study population;
4. To investigate rural/urban living and dementia outcome using Cox Proportional Hazard (CPH) models, adjusting for confounders (childhood intelligence, and deprivation).

I now discuss my study in the context of Scotland (including its population, geography and dementia prevalence), and the SMS1947 with an overview of these aspects in the next section.

5.3 Dementia and the Scottish Mental Survey 1947

5.3.1 An overview of Scotland's population, healthcare and geography

Scotland has a population of approximately 5,436,600 people as of 2022 [662]. The distribution between rural and urban areas shows that approximately 17% of the population live rurally, whilst 98% of Scotland's land mass is rural [338, 662]. Interestingly, the rural population of Scotland is growing at a faster rate than the rest of Scotland, mainly due to inward migration and increased accessibility to rural areas [338].

Scotland can be categorised geographically by a number of different scales. These

include: postcode sector, census output areas, data zones and the more recent development of the Urban-Rural Classification 2020 system [663]. The latter (the Urban-Rural Classification 2020 system) has provided a reliable method of mapping the geographical landscape comparing layers of rurality or urbanicity across Scotland and can be valuable for research and official reports [663]. It was created in Scotland incorporating two main criteria: population and accessibility [663]. It is meaningfully applied to small area geographical areas in Scotland.

Other small area geographical areas include the census output areas, which are the smallest scale geographically in the Census (undertaken every 10 years in the UK) that can be used for census statistics. Data zones are made up of aggregates of census output areas. Both census output areas and data zones are relatively stable geography markers as they change every ten years following the UK Census [663].

I have outlined a comparison of the small geographical scales in Scotland in Table 5.1 below and will subsequently discuss why I selected the Urban-Rural Classification 2020 system for my final study project, after provisionally planning to use data zones and postcode sectors initially.

Scale	Started	Description
Postcode sector [664]	1932	A postcode is called a ‘postcode unit’ in full. A postcode sector comprises 4 digits of a Scottish postcode (postcode unit). Postcodes are identifiable variables in healthcare research. Postcode sectors are less likely to be identifiable if investigating a common disease with high prevalence in an area, but still have theoretical risks of this.
Census Output Areas (Output Areas, OAs) [665]	2001	Smallest geographical scale for census statistics. Change depending on population size at census (every 10 years). Represents 40-250 households (100-625 individuals). Grouped together become Lower Layer Super Output Areas (LSOAs) or Middle Layer Super Output Areas (MSOAs).
Data zones [665]	2001	6,976 data zones in Scotland. Aggregates of OAs. Small scale but large enough for non-disclosure statistics. 500-1,000 household residents of similar social characteristics. Largely aligned with physical boundaries (e.g. councils). Can be grouped to represent larger areas.
Urban-Rural Classification system [663]	2020	Developed by the Scottish Government. Postcodes, census output areas and data zones used to create the system. The classification system is available in levels called ‘fold’. Areas are classed by their level or urbanicity or rurality from population size and accessibility.

Table 5.1: Geographical scales in Scotland

For larger geographical areas, the Urban-Rural Classification 2020 system can be used to “*show the degree of urbanity, rurality and remoteness as a proportion of the population living within each of the 6- and 8-fold categories*” [663]. The Urban Rural classification is available as 2-,3-,6- and 8-fold categories [663]. The 6-fold categories are: Large Urban Areas, Other Urban Areas, Accessible Small Towns, Remote Small Towns, Accessible Rural and Remote Rural [663]. In comparison,

the 8-fold categories are: Large Urban, Other Urban, Accessible Small Towns, Remote Small Towns, Very Remote Small Towns, Accessible Rural, Remote Rural and Very Remote Rural [663].

Figure 5.1 shows a map of Scotland with the variation in rural and urban areas based on this classification taken from [663]. I used the 2003-2004 version of this classification for my research as a measure of geographical variation (urban/rural). This version is not specific for the year 2003-2004, but the classification from that year (2003-2004) has been used to assess other years so there is consistency in using one year's classification system across many years [663].

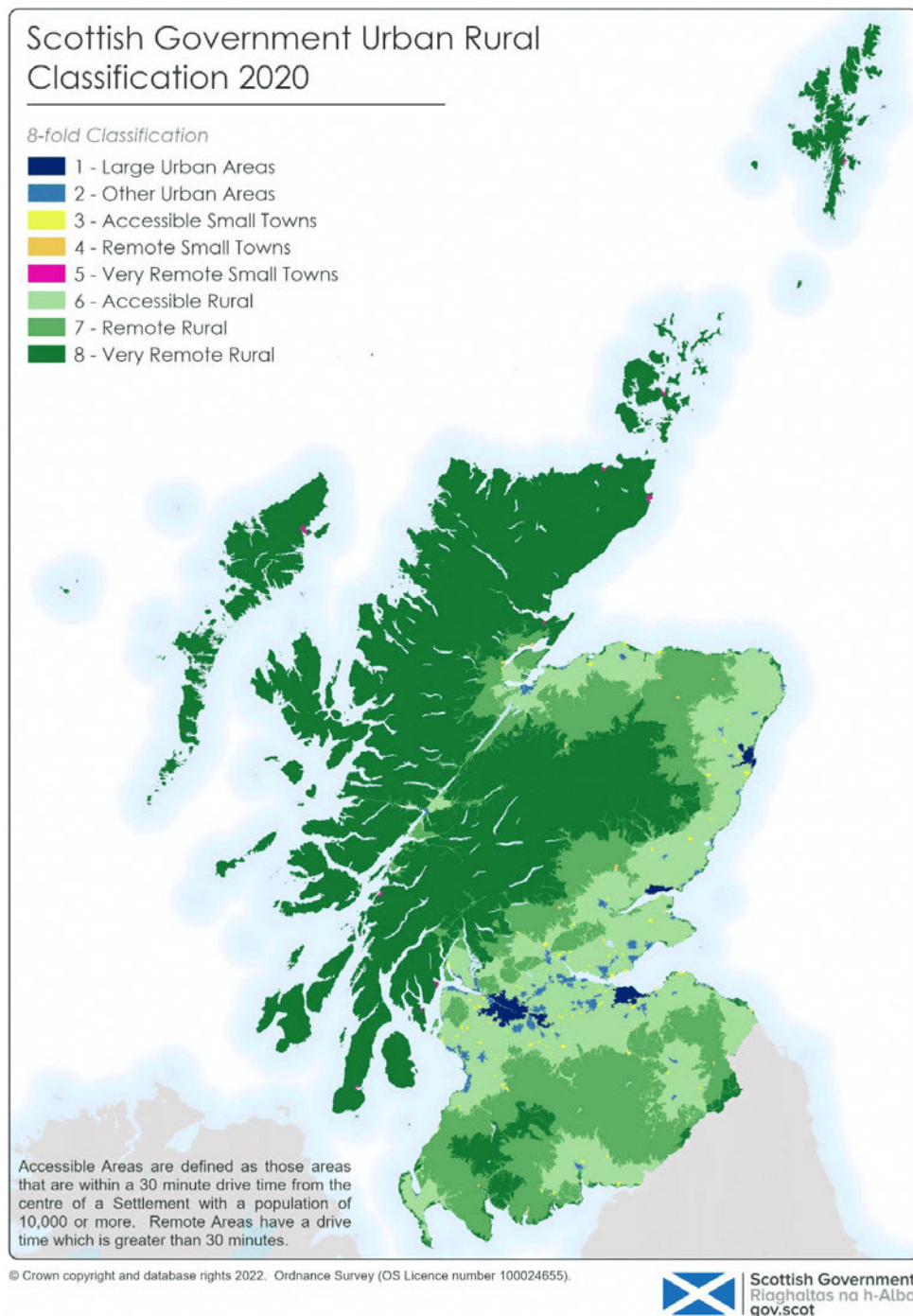


Figure 5.1: Map of the Scottish Government 8-fold Urban-Rural Classification 2020 system [663]

Source: Scottish Government 8-fold Urban Rural Classification 2020 map. Author: Scottish Government, 2020. Used with permission under education licence from ©Crown copyright and database rights 2022, Ordnance Survey (OS Licence number 100024655), under the Open Government Licence (OGL):
<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>.

Scotland also has unique deprivation indexes: the Scottish Index of Multiple Deprivation (SIMD), and the Carstairs Index. What constitutes a geographical area can also influence other indexes such as deprivation scales. Table 5.2 below highlights descriptions of the two deprivations scales.

Scale	Started	Description
Carstairs Index [666]	1980s	Also called The Carstairs and Morris Index. Uses postcode sector from Census data and also a geographical calculation of Consistent Areas through Time (CATTs). Comprises four indicators representing material disadvantage: lack of car ownership, low occupational social class, overcrowded households and male unemployment. Last revision was 2011. It does not use health indicators.
Scottish Index of Multiple Deprivation (SIMD) [667]	2004	A Scottish Government standardised approach to identify geographical areas of multiple deprivation across Scotland. Holistic approach. Created from 6,976 data zones. Seven domains assessed including income, employment, education, health, access to services, crime and housing. Each domain is ranked and then combined. Overall SIMD provides a measure of relative deprivation at data zone level. SIMD can indicate one data zone more deprived than another but not by how much. Commonly areas highlighted below a certain rank (e.g. the 5%, 10% etc most deprived data zones in Scotland).

Table 5.2: Deprivation scales in Scotland

Of importance, the SIMD can be less helpful in rural areas in capturing deprivation and often reflects a more mixed picture [\[667, 668\]](#). This is because the data zones that comprise the small area assessed in the SIMD are larger in rural areas. As they cover a larger land area, they can reflect a more mixed picture with varying levels of deprivation. Due to this, the SIMD are less able to capture

small pockets of deprivation within rural areas compared with the larger pockets in urban areas. The Cartairs Index is not felt to be a good reflection of rural deprivation with its indicators [668, 667].

One further important aspect of Scotland relevant to my study is the national healthcare system (NHS), and the Community Hospital Index (CHI) number [482]. When considering geographical variation and health outcomes in Scotland, the CHI is essential. This was discussed in Section 3.5.1 (Unique identifiers) and is a unique number given to every individual in Scotland, consisting of their date of birth and a unique identifier. It is with the individual from birth to death.

Scotland is well placed to study dementia using routinely collected health data with its excellent potential for data linkage at a national level due to the population-wide unique identifier, the CHI-number. As mentioned in Chapter 3, Section 3.5 (Data linkage: unique identifiers, types of linkage, datasets), having a unique identifier aids the ability to link healthcare data. The CHI-number system was developed in the 1980s [483]. It is a ten digit number which includes a person's date of birth at the start in the format of two digits for day/month/year and followed by a numerical code for sex at birth, and finally a four digit number (a unique identifier) at the end. As it includes the individual's date of birth, it is a more accurate, and therefore powerful, unique identifier than the NHS number. Whilst the NHS number is a unique identifier there is, as outlined in Chapter 3, the potential for a person attending a different NHS hospital to be issued a new NHS number if their original one cannot be located immediately.

Using the unique identifier that each individual has, data can be linked with routinely collected health data and made available for researchers in an anonymised and de-identified format within the National Safe Haven. Routinely collected health data available in Scotland include: the Scottish Morbidity Records (SMR), the Prescribing Information System (PIS) and the National Records of Scotland Mortality datasets (NRS mortality/NRS deaths) [669]. I outline the details of

these routinely collected health data that I used, including a description of the data they cover in more detail, in the study design Section [5.5.1](#) discussing the datasets in the context of dementia ascertainment.

5.3.2 Dementia in Scotland

In 2024 it is estimated that there are approximately 90,000 people living with dementia in Scotland, with approximately 97% over the age of 65 years [\[101\]](#). The Scottish Government anticipate a 50% increase in people with dementia over the age of 65 years in the next 20 years [\[101\]](#). In 2023, the Scottish Government released a new Dementia Strategy for Scotland [\[96\]](#). This ten year vision will have the first delivery plan implemented for 2024-2025. It builds on from the past National Dementia Strategy, and the first report from Information Services Division (ISD) Scotland/NHS Scotland in 2016 highlighting estimated, and projected dementia diagnosis rates, in Scotland from 2014 to 2020, based on data from Scottish health boards [\[670\]](#).

In the 2016 report, it was recognised that the fear of dementia means people delay coming forward for a diagnosis. This context is important to consider in relation to a changing, and growing, national drive to encourage a diagnosis of dementia and how this might impact on interpreting results over time such as age-standardised prevalence rates. I discuss this further in my results in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) and in the overall thesis discussion in [Chapter 7: Discussion](#). There are reports on factors influencing individuals seeking a diagnosis of dementia worldwide [\[128, 671, 672, 647, 131\]](#). This is important to consider in the context of ‘diagnostic gap’ in dementia diagnosis and how this might influence dementia ascertainment in my research in this chapter.

Longitudinal studies, particularly at a population level, can be valuable for investigating dementia. For my research question, it is particularly helpful to have a longitudinal perspective when considering the impact of a rural or urban lo-

cation and dementia rates. The Scottish Mental Survey 1947 (SMS1947) linked to routinely collected health data can provide a valuable population level, with individual level data, longitudinal study to investigate my research question.

5.3.3 The Scottish Mental Survey 1947

The SMS1947, is a historical population-level survey which included individuals born in Scotland between 1935-1937, whom at the age of eleven years undertook a national childhood intelligence test, called the the Moray House Test (MHT), on 4th June 1947 across the whole of Scotland [673]. The test was undertaken with the same instructions and within a 45-minute test period on the same day [673]. The Moray House Test version no.12 was used, which is an intelligence test, and from this an intelligence score was derived and corrected for age in days at the time of testing. This amounted to approximately 70,000 individuals with intelligence results at age eleven [673]. Additionally, some specific groups of children who were twins or selected as having siblings had further assessments undertaken [674, 673]. For this research I will refer to the MHT and intelligence test interchangeably.

Sub-groups from the SMS1947 have been subsequently followed up in the north of Scotland, Aberdeen (Aberdeen Birth Cohort ABC1936) and in the South/Central Scotland, Lothian (Lothian Birth Cohort LBC1936) [674, 502, 675, 673]. These formed nested cohort studies, of which participants were also part of the SMS1947 in the Aberdeen or Lothian areas. These nested cohort studies have allowed more detailed longitudinal and face-to-face investigations to be undertaken [673]. Those who were part of the SMS1947, and still alive, would be aged around 88 years old now, an ideal age for investigating dementia given the prevalence of late onset dementia rises with age and is expected to be one in six by age 80 [670].

A major strength of using this historical survey (SMS1947) in dementia research is the information on childhood cognition at an individual-level but population-scale from 1947 [673]. Higher childhood, or early adulthood, intelligence has

been shown in several different countries to be associated with a lower risk of mortality by mid to later adulthood [676, 677, 678, 679, 680]. It is unclear why this association exists and seems to be independent of socio-economic status of origin or perinatal influences.

Additionally, it has been shown that childhood intelligence can influence later dementia risk, and other morbidities and mortalities [676, 677, 678, 681]. The risk of dementia related mortality was reduced by approximately a third in those with higher intelligence scores [681]. Furthermore, this difference varied between men and women [681]. In men there was a 10%, and in women 24%, reduced risk of dementia for one standard deviation (SD) higher score in childhood intelligence [681]. It is an important confounder to adjust for in dementia studies.

In summary, this section highlights why the historical survey (SMS1947) provided valuable information for investigating dementia outcome when linked to routinely collected health data to investigate my research question.

5.3.3.1 Data linkage of Scottish Mental Survey 1947

The SMS1947 on its own consisted of hand-written ledgers, which decades later, when discovered, were subsequently digitalised. Given the detailed information contained in the SMS1947, researchers recognised that linkage of this population-wide survey to routinely collected health data would be valuable for research [673, 682]. The SMS1947 has been linked to routinely collected health data for two large research projects. Project One was entitled the *‘Lifelong health and well-being of the ‘Scotland in Miniature’ cohort: The 6-Day Sample of the Scottish Mental Survey 1947’*. Project Two was entitled the *‘Mental health within the family and between generations–Phase 1: Linking the Scottish Mental Survey 1947 cohort to mental health outcomes’*.

Project One had CHI-linkage of the SMS1947 participants to routinely collected health data until their 70s. This project was near the approved end date for ethics

and permissions when I received approval to use this project for my research question. The second project is still fully active, with data linkage up until 2018 (participants were in their 80s). Over the first two years of my PhD it became more appropriate to use Project Two for my research question.

For this section, I will refer to the Project One above as the ‘*Scotland in Miniature*’ study, and Project Two as the ‘*Mental health outcomes in SMS1947*’ study. My research (results presented in [Chapter 6: Results of the SMS1947 dementia data linkage study](#)) is part of the second Project ‘*Mental health outcomes in SMS1947*’ study. I have included the processes I went through in Table 5.6 ([Timeline of planning SMS1947 research in Scotland](#)) for Project One, and then subsequently Project Two as it highlights an important learning aspect of my PhD in streamlining timely access to CHI-linked routinely collected health data.

Project Two which I used for my research looked at all mental health outcomes. As dementia is a key mental health outcome, my study was well placed within this project. My study was the first investigating dementia as one of the mental health outcomes in the SMS1947 linked to routinely collected health data. For clarity I will now refer to my study as the ‘*SMS1947 dementia data linkage study*’.

The data linkage of the SMS1947 to routinely collected health data was established prior to my research. It was complex, and involved substantial work in tracing individuals virtually and linking in several stages by data analysts [682]. Figure 5.2 below shows the multi-step process undertaken prior to my research to enable all individuals within the SMS1947 (and traced) to be linked to their corresponding CHI-number (an individual’s unique identifier in health data in Scotland). The CHI-number started in the 1980s in Scotland, so data scientists sought to link individuals from the SMS1947 to their corresponding subsequent CHI-number. A ‘Master Index Key’ is used to link all datasets. This allows linkage with a ‘key’ that de-identifies individuals. This is also outlined in Figure 5.2 below. The researcher does not have access to the CHI-numbers ensuring the

‘safe data’ component of the Five-Safes Framework (Table 3.2) is in line with data governance.

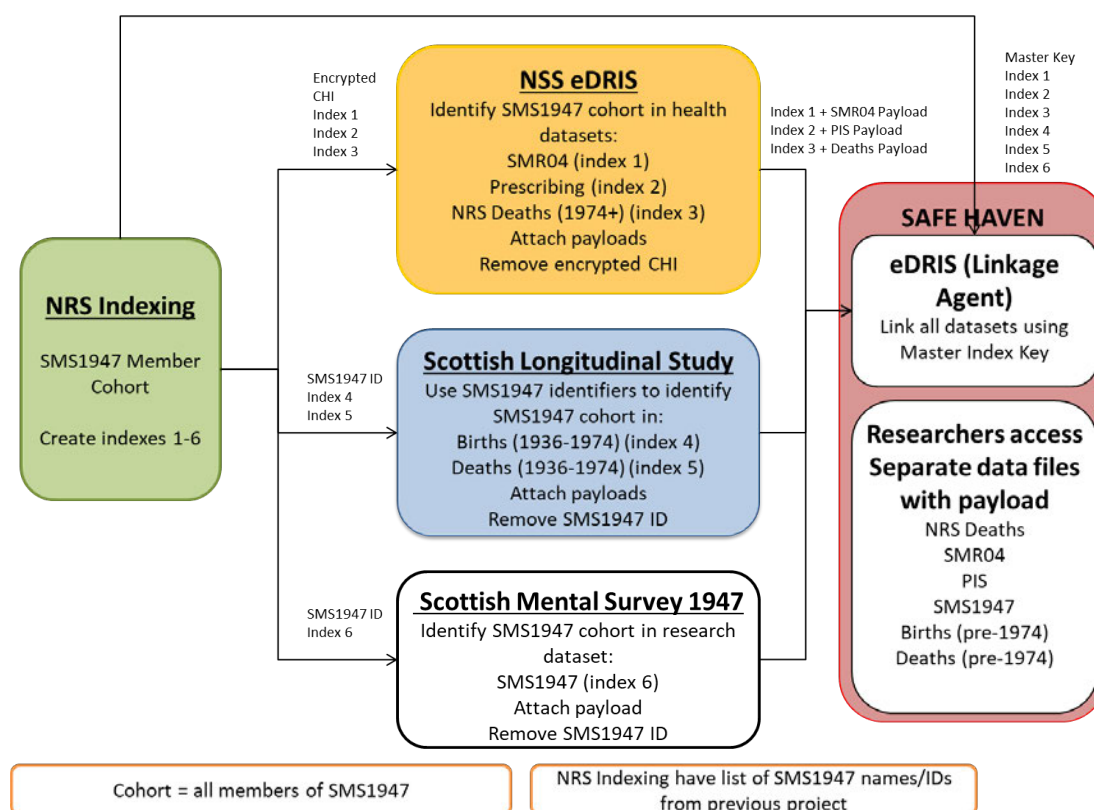


Figure 5.2: Data linkage procedure for the SMS1947 Project Two

Source: From the Public Benefit and Privacy Panel for Health and Social Care application entitled '1819-0073 Deary HSC-PBPP Application' for the SMS1947 research project 'Mental health within the family and between generations – Phase 1: Linking the Scottish Mental Survey 1947 cohort to mental health outcomes' referred to as Project Two in this thesis.

Authors of Figure: Matthew Iveson and Amy Tilbrook. 2018. Used with author permission.

Data linkage of the SMS1947 to routinely collected health data resulted in $n=53,037$ individuals (***database*** population) of the $n=70,805$ individuals in the SMS1947 (***source*** population) being linked and available in Scotland's National Safe Haven for my research study. I accessed these data ($n=53,037$), checked and cleaned the data, created the ***study*** population from this ***database*** population, designed and performed the analysis (Section 5.5.2).

In summary, in Section 5.3 ([Dementia and the Scottish Mental Survey 1947](#)) I have given an overview of Scotland in this section with particular focus on its different geographical and deprivation scales; healthcare system and CHI number; dementia prevalence in Scotland; and the Scottish Mental Survey 1947 (including its data linkage to routinely collected health datasets). I present the geographical scale and deprivation scale I used for my *SMS1947 dementia data linkage study* in the following sub-sections.

Geographical scale - SMS dementia data linkage study

The geographical scale I selected for my research was the Urban-Rural Classification 2020 system version 2003-2004 year [\[663\]](#). Initially, I wanted to investigate several layers of geographical scale (postcode sector, data zones and census output areas) in Project One ‘*Scotland in Miniature*’, however as I moved to align my study with Project Two ‘*Mental health outcomes in SMS1947*’, these variables were no longer available in Project Two.

The development of the Urban-Rural Classification 2020 classification system [\[663\]](#) in Scotland provided a unique opportunity to focus my research question on rural and urban areas within the *SMS1947 dementia data linkage study*. I used one classification of Urban-Rural Classification system (the 2003-2004 version) for the analysis so I did not introduce bias using multiple different versions. I selected the 8-fold version for greatest range of urban and rural classes. The 2003-2004 version was selected as it was felt to be the most consistent over time.

Deprivation scale - SMS dementia data linkage study

The deprivation scale I selected for my SMS1947 dementia linkage study was SIMD 2004 [\[667\]](#). The reasons for selecting the SIMD over the Carstairs was the recommendation from Public Health Scotland to use SIMD for research after 2011 and, that it is perceived as a better reflection of deprivation [\[668\]](#). Additionally, I selected this specific year of SIMD (SIMD 2004) because it was closest in

chronological age to the Urban-Rural 8-fold Classification 2003-2004 used. This introduced less bias.

There are a number of SIMD variables: SIMD decile, SIMD standardised, SIMD ordered [667]. The SIMD decile variable is the raw data [667]. The SIMD standardised version treats SIMD as a continuous variable (i.e. only includes a linear effect), but standardised (i.e. mean= 0, SD= 1) [667]. The SIMD ordered variable treats SIMD as an ordered categorical, so including it will estimate linear, quadratic, cubic effects [667]. As the SIMD decile variable was the raw data, I used this for adjusting in my Cox Proportional Hazard models outlined in Section 5.5.5 (Statistical analysis) of this chapter.

Project Two, ‘*Mental health outcomes in SMS1947*’ provided a powerful population-wide survey from 1947 with individual-level data, including childhood intelligence. This allowed childhood intelligence to be adjusted for as a confounder in the analysis. This historical survey (SMS1947) which was CHI-linked had 2018 as the censorship year. This means the participants were traced and CHI-linked to routinely collected health data up until (and including) 2018. At that point, individuals were classed as ‘alive’ if no evidence of death at that point in time. Those alive were in their 82nd year at the end of censorship, an important age group for ascertaining dementia in.

I used the MHT (intelligence score) for each individual in the SMS1947 as a marker of childhood intelligence. There are several different versions of the MHT variable within the National Safe Haven (NSH). I used the variable ‘MHT_corr_std’ which stood for the MHT (intelligence score) standardised to a mean=0 and SD= 1 for all SMS1947 participants. There is recognition that the MHT (intelligence score) at eleven years old could be linked with deprivation as those in higher areas of deprivation may not have the same education attainment as people living in lower areas of deprivation. Different scores for the MHT (intelligence score), might be reflective of socio-economic factors. This could be considered collinear-

ity and therefore I checked for this in my analysis.

5.4 Data access and permissions processes

Prior to discussing my study design, I outline the data access and permissions process in Scotland for routinely collected health data in this section. This is important to give context to my study design and subsequently the timelines I discuss at the end of this chapter as important learning from this process. In Chapter 3 I highlighted the Five-Safes Framework (Table 3.2). In this section I outline the process in Scotland for data access and permissions in relation to this Five-Safes Framework.

The Five-Safes	SMS dementia linkage study
Safe data	No researcher has access to the CHI-number/name or identifiable details. There is a two step process in ensuring data linkage between the SMS1947 and routinely collected health data takes place in stages to ensure anonymity. This was previously undertaken by data analysts at Public Health Scotland, not a researcher.
Safe projects	Approval from Scotland's Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) was required for all projects using routinely collected health data. Amendments required further approval and discussion with Public Health Scotland eDRIS research coordinator prior to submission.
Safe people	I underwent numerous pre-defined initial training and subsequent renewal as outlined in Table 5.5 (Timeline of training undertaken to access data in National Safe Haven)
Safe settings	Scotland's National Safe Haven (NSH) was the Trusted Research Environment (TRE) which was a SecureLab environment with a two login/password protection for only approved researchers. It could be accessed remotely through strict controls.
Safe outputs	Public Health Scotland eDRIS research coordinator screened all my results and formally approved them as non-disclosing outputs to be used in this thesis

Table 5.3: The Five-Safes Framework for the *SMS dementia data linkage study*

Data governance was controlled by a number of different organisations. All needed notified when any minor or major amendments were planned. Over the course of my PhD, the governance bodies have changed slightly. Currently, in Scotland, access to routinely collected health data is now managed by Public Health Scotland (PHS) electronic Data Research and Innovation Service (eDRIS) [542]. This replaced the prior management by NHS Scotland's Information Service Division

(ISD) overseeing eDRIS [683]. There are several stages to allow data linkage to take place and then allow researchers to access the linked data.

In addition to IRAS (Integrated Research Application System) [514] that covers the UK as a system for applying for health and social care research in the UK and the local ethics committees in Scotland, there are a number of key governing bodies in Scotland. Table 5.4 outlines these, and their remit, for giving access and permission to using routinely collected data for my research proposal. There were additional governing bodies during the course of my PhD that are no longer active. This reflects the changing nature of health data research, and aims to streamline the permissions, approvals and accesses in lines with the Fives-Safes Framework. The University of Edinburgh (UoE) also plays an important role in governance as the sponsor of the research and lead organisation for the research study. Any changes must also be signed off by the UoE data protection officer assigned to the study, prior to submission to ethics or HSC-PBPP [452]. UoE is also the Scotland site for the UK Health Data Research (HDR) so coordinates with the other Scottish Universities in this role.

Organisation	Started	Description
Electronic Data Research and Innovation Service (eDRIS) [542]	2013	Established to support the use of health data and electronic medical records in NHS Scotland institutions. eDRIS is a portal to national data and the National Safe Haven. Works in conjunction/collaboration with local safe havens in partnership with main Scottish academic institutions. The role includes facilitating access to datasets; enabling data linkage; and secure data transfer of data within Scotland's National Safe Haven in collaboration with Edinburgh Parallel Computing Centre (EPCC) and NHS Scotland. A researcher is allocated an eDRIS research co-coordinator to discuss new research, highlight key issues and review applications prior to submitting to HSC-PBPP. They also review any research amendments and ensure researchers have appropriate GDPR training. They provide a link between the data linkage team and researchers for questions to be clarified.
Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) [452]	2015	Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) is a governance structure for NHS Scotland for research in Scotland. It scrutinises and considers a researcher's application for the purpose of benefit for the public and for information governance requirements. The HSC-PBPP ensures the researcher's request to access data meets and maintains the Information Governance principles of the Five-Safes Framework. There are two tiers of application process. Applications only proceed to tier 2 approval if the tier 1 committee require it. It succeeded both the NHS National Services Scotland's Privacy Advisory Committee (PAC) and the National Caldicott Scrutiny Panel (NC).
Public Health Scotland (PHS)	2020	Public Health Scotland is an NHS board sponsored by the Scottish Government and the Convention of Scottish Local Authorities (COSLA). It is a large organisation and not solely focused on research. Other services linked with research were moved into this organisational governance structure during my PhD. Within PHS, eDRIS is situated and the CHI-Indexing and Linkage team (CHILI) [482], which is part of the Data and Digital Innovation directorate of PHS.

Table 5.4: Organisations and governance structures in Scotland

The first step in developing a research study was to request a meeting with the eDRIS research coordinator to discuss my study plans and how it fitted within the overall aims of the SMS1947 CHI-linked data research project in the National Safe Haven. Once the eDRIS researcher confirms a new research study is a ‘safe’ project for public benefit, and still meeting the overall aims of the linked SMS1947 study in the National Safe Haven, then the next stage is to submit an application to Scotland’s Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) [452].

HSC-PBPP must approve access to the data for a research study. For my first project plan I had to complete a full application form and major amendments. In Table 5.6 ([Timeline of planning SMS1947 research in Scotland](#)), I outline these multiple steps required for both Project One, and subsequently Project 2 (used for my research question in [Chapter 6: Results of the SMS1947 dementia data linkage study](#)). An important part of Table 5.6 is the extensive timeline for amendments for Project One, versus Project Two.

The process for the HSC-PBPP approving a researcher’s application is highlighted in Figure 5.3 [452]. The eDRIS research co-coordinator submits the application, completed by the researcher, to the HSC-PBPP on the researcher’s behalf.

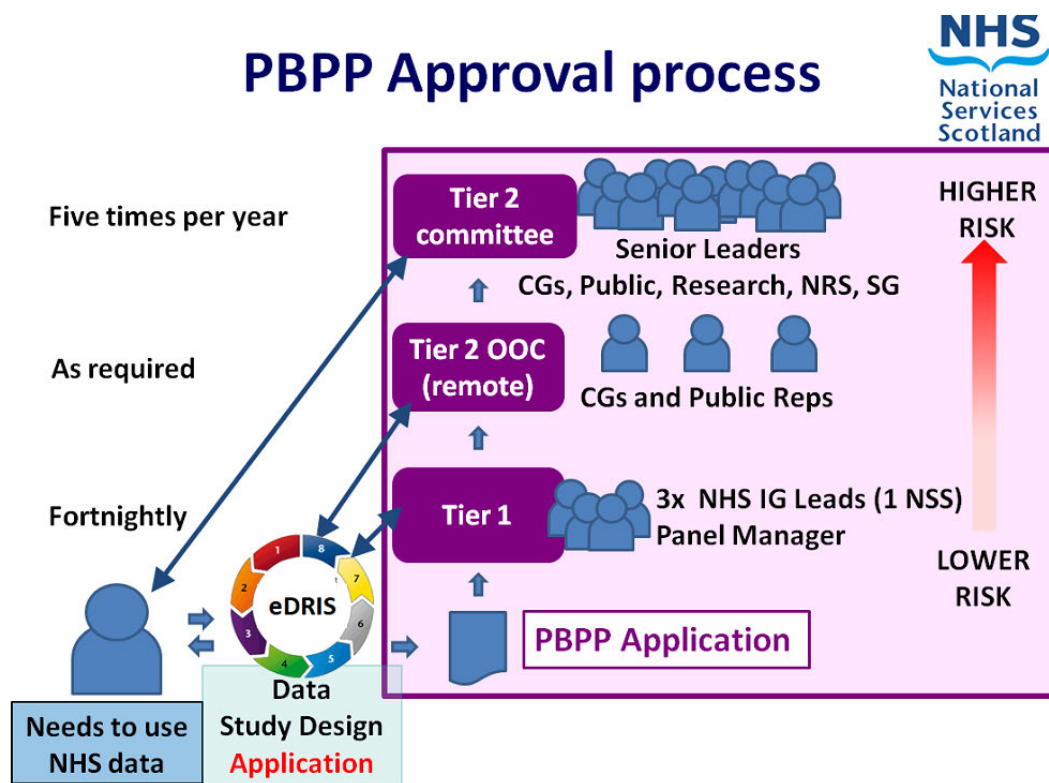


Figure 5.3: The HSC-PBPP approval process in Scotland [684]

Source: Public Health Scotland, Public Benefits and Privacy Panel for Health and Social Care.

Used with permission via Open Government Licence (OGL) v3.0, available at:

<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>.

Once the research has approval from the HSC-PBPP [452] (and ethics is already in place), then training must be undertaken prior to allowing a researcher access to the data within a Trusted Research Environment (TRE): which for Scotland is the National Safe Haven (NSH). This is in keeping with ‘safe people’ and ‘safe place’ as part of The Five-Safes Framework (Table 5.3). In Table 5.5 I outline the training I undertook as a researcher to meet the requirements.

Date	Course
16th Dec 2015 - 13th Feb 2016	Health & Social Care Information Centre (hscic): • Secure Handling of Confidential Information • Role of Caldicott • Patient confidentiality • Caldicott Guardian in NHS & Social care • Introduction to Information Governance • Information Governance: The Beginner's Guide • Access to Information & Information sharing • Password management • Information security guidelines • Secure transfers of personal data • Information security management
5th Jul 2016	NHS National Institute for Health and Research (Introduction to Good Clinical Practice eLearning, Secondary Care): • Introduction to Research in the NHS • Good Clinical Practice and Standards in Research • Study Set Up and Responsibilities • The Process of Informed Consent • Data Collection and Documentation • Safety Reporting
13th May 2019	MRC: Research, GDPR and confidentiality
29th Jan 2021	Office for National Statistics (ONS) Safe researcher training for the Scottish Centre for Administrative Data Research (SCADR) course
1st Mar 2021	ONS SCADR examination
9th Jan 2024	UKRI and MRC: Research, GDPR and confidentiality

Table 5.5: Timeline of training undertaken to access data in National Safe Haven

5.4.1 Timeline for designing the SMS1947 dementia data linkage study

I have outlined the timelines for designing my study (the *SMS1947 dementia data linkage study*) using Project Two (*‘Mental Health Outcomes SMS1947’*) which this Chapter and the following [Chapter 6: Results of the SMS1947 dementia data linkage study](#) refers to in Table 5.6 ([Timeline of planning SMS1947 research in Scotland](#)). Additionally, I have included the initial steps for Project One (*‘Scotland in Miniature’*) in the Table 5.6 ([Timeline of planning SMS1947 research in Scotland](#)) to highlight how much time can be required to access routinely collected health data with a *nested* historical survey, even when a project has previously been through the process of data linkage. This has been important learning as part of the PhD training. I will focus this chapter on the methods used for the research results in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) to avoid confusion.

Date	Project No.	Process	Progress	Body	Information
25.08.2015	1	Non-substantial amendment	<i>Submitted</i>	Scotland Regional Ethics Committee (REC) A	To add me as a researcher to the ‘ <i>Scotland in Miniature</i> ’ study submitted to Scotland REC A.
07.09.2015	1	Non-substantial amendment	<i>Submitted</i>	NHS Lothian Research & Development (R&D)	To add me as a researcher to the ‘ <i>Scotland in Miniature</i> ’ study submitted to NHS Lothian R&D.
08.09.2015	1	Non-substantial amendment	<i>Approved</i>	Scotland REC A	Approval of me as a researcher on the ‘ <i>Scotland in Miniature</i> ’ study.
08.09.2015	1	Non-substantial amendment	<i>Approved</i>	NHS Lothian R&D	Approval of me as a researcher on the ‘ <i>Scotland in Miniature</i> ’ study.
16.11.2015	1	Substantial amendment	<i>Submitted</i>	IRAS	Amendment 4. Addition of my research project proposal to investigate geographical variation of dementia in the SMS 1947 within the ‘ <i>Scotland in Miniature</i> ’ study.

Continued on next page

Date	Project No.	Process	Progress	Body	Information
25.11.2015	1	Substantial amendment	<i>Submitted</i>	NHS Lothian (R&D)	Submission to NHS Lothian R&D for amendment number 4 to include my revised research project within the research protocol.
23.12.2015	1	Substantial amendment	<i>Approved</i>	Scotland REC A	Approval from Scotland REC A committee for substantial amendment number 4 if PBPP also approves.
24.12.2015	1	Substantial amendment	<i>Approved</i>	NHS Lothian R&D	Approval from NHS Lothian R& D for above substantial amendment number 4 if PBPP also approves.
02.03.2016	1	Substantial amendment	<i>Submitted</i>	HSC-PBPP	Submission of Substantial amendment PBPP form with appendices and including approval letters from ethics.
01.04.2016	1	Substantial amendment	<i>Approved</i>	HSC-PBPP	Approval granted: reference 1516-0373/Walesby-Geographical variation in dementia in the 1947 Scottish Mental Survey population - an extension to the original ' <i>Scotland in Miniature</i> ' study.
21.04.2016	1	Notification from eDRIS		eDRIS	Request to read, agree and sign the National Services Scotland (NSS) eDRIS User Agreement to access the National Safe Haven
26.06.2017	1	Non-substantial amendment	<i>Submitted</i>	HSC-PBPP	Addition of researchers: my supervisor (Dr Tom Russ) and two further researchers to assist with cleaning the data

Continued on next page

Date	Project No.	Process	Progress	Body	Information
29.06.2017	1	Non-substantial amendment	<i>Approved</i>	HSC-PBPP	Approval of above additional researchers, research must also now notify all relevant bodies involved in the project.
29.06.2017	1	Notification from eDRIS		eDRIS	All new additional researchers to sign eDRIS user agreement and for me to submit these to the eDRIS research coordinator.
01.12.2018 to 30.03.2019	1	Multiple discussions		Multiple	Discussions regarding changing Chief Investigator following the sudden death of Professor John Starr, with a number of organisations (University of Edinburgh, as research sponsor, and also co-PI Prof Ian Deary and future Chief Investigator, Dr Tom Russ). Chief Investigator must be NHS based to comply with requirements of IRAS. Challenges with accessing IRAS with the login for the Chief Investigator via individual email/password. Formal approval from IRAS to transfer full study temporarily to Dr Tom Russ to create the amendment and submit as new Chief Investigator but complicated. Further request to IRAS for a permanent transfer of the study to a different IRAS account (other than Prof Starr's) requiring a number of separate steps.

Continued on next page

Date	Project No.	Process	Progress	Body	Information
18.02.2019	1	Substantial amendment	<i>Submitted</i>	Scotland REC A	Amendment number 6 (Chief Investigator to Dr Tom Russ after the sudden recent death of Professor John Starr, previously Chief Investigator). Submission included amended IRAS NHS REC form, revised research protocol and cover letter.
29.03.2019	1	Substantial amendment	<i>Approved</i>	Scotland REC A	Amendment number 6 approved, still requires approval from HSC-PBPP.
04.04.2019	1	Non-substantial amendment	<i>Submitted</i>	NHS Lothian R&D	Extension to the project. A request to refresh the already linked data the study until December 2018 so the results and analysis (and subsequent publications) are up to date. The previous data was until December 2015.
05.04.2019	1	Substantial amendment	<i>Approved</i>	NHS Lothian R&D	Amendment 6 approved
08.04.2019	1	Non-substantial amendment	<i>Approved</i>	NHS Lothian R&D	Approval of non-substantial amendment for the extension of the project to December 2018. Researcher to notify all relevant bodies involved in the project.

Continued on next page

Date	Project No.	Process	Progress	Body	Information
10.05.2019	1	Substantial amendment	<i>Submitted</i>	HSC-PBPP	Amendment 6. Submission to HSC-PBPP of revised protocol and amendment for three changes: 1) An updated (refresh) of the data to December 2018, 2) Change of Chief Investigator, 3) Remove researchers no longer requiring access to the project
2019	1	Substantial amendment	<i>Approved</i>	HSC-PBPP	Approved Amendment 6 changes: 1) an updated (refresh) of the data to December 2018, 2) Change of Chief Investigator, 3) Remove researchers no longer requiring access to the project.
12.05.2019	1	Non-substantial amendment	<i>Submitted</i>	NHS Lothian R&D	Addition of researcher, Dr Susan Shenkin as my new supervisor since death of previous supervisor Professor Starr.
15.05.2019	1	Non-substantial amendment	<i>Approved</i>	NHS Lothian R&D	Approved addition of Dr Susan Shenkin as additional researcher. I need to notify all relevant bodies involved in the project.
03.03.2021	2	Non-substantial amendment	<i>Submitted</i>	Public Health Scotland (PHS)	Addition of me as a researcher to the second project.
11.03.2021	2	Non-substantial amendment	<i>Approved</i>	PHS	Approval of me as a researcher to the second project.

Continued on next page

Date	Project No.	Process	Progress	Body	Information
Oct 2023	2	Discussions		eDRIS	Discussions about ICD codes for dementia and flags for dementia-related ICD codes in the deaths dataset. Current approval for other mental health ICD codes but not dementia ICD codes so agreed that a non-substantial amendment was required to access this information.
16.10.2023	2	Non-substantial	<i>Submitted</i>	HSC-PBPP	Submitted to HSC-PBPP request for NRS deaths data with ICD codes for dementia and flag to ascertain dementia from deaths datasets.
31.10.2023	2	Non-substantial	<i>Approved</i>	HSC-PBPP	HSC-PBPP approved the request for NRS deaths data with ICD codes for dementia and flag to ascertain dementia from deaths datasets.
Nov 2023	2	Discussions		eDRIS	Discussions with eDRIS research coordinator to extend project end date and add additional predictors.
06.12.2023	2	Non-substantial amendment	<i>Submitted</i>	HSC-PBPP	Submitted by co-researcher. Request to 1) extend research end date to 31st January 2025 and 2)add additional predictors (geographical data, Urban-Rural 8-fold 2003-2004 classification) to PIS, NRS Deaths datasets and SIMD/Carstairs.
15.12.2023	2	Non-substantial amendment	<i>Approved</i>	HSC-PBPP	Approval to extend project date and add additional geographical predictors

Continued on next page

Date	Project No.	Process	Progress	Body	Information
25.04.2024	2	Notification from eDRIS	Data linkage	eDRIS	Additional predictors (geographical data, Urban-Rural 8-fold 2003-2004, added to datasets missing this (PIS and NRS mortality) and also deprivation predictors SIMD/Carstairs. Available to use for research.

Table 5.6: Timeline of planning SMS1947 research in Scotland

5.5 Study design of the SMS1947 dementia data linkage study

This was an observational retrospective study using a population-based historical survey (SMS1947) linked to routinely collected health data. The data was only accessible via the National Safe Haven (NSH) in Scotland, which is a secure point of access known as a Trusted Research Environment (TRE) [669]. The SMS1947 database was previously linked by data scientists in eDRIS to anonymised records from: 1) national administrative databases [Birth records (1935-1937, Scottish Longitudinal Study, SLS), Death Records (Deaths 1975-present, NRS, and Deaths 1935-1975 (SLS),] and 2) health databases: the Scottish Morbidity Records (SMR) 01 (general admissions) and 04 (psychiatric admissions), and the Prescribing Information System (PIS) data [669].

To ensure anonymity, 1) all researchers were unable to see CHI-numbers (this was used only for the data linkage by the data analysts), 2) age for all individuals available in months rather than date of birth to ensure anonymity and 3) data could only be accessed and viewed within the secure Scottish NSH.

5.5.1 Datasets used for dementia ascertainment

The datasets available within the NSH relevant for my research were the Scottish Morbidity Records (SMR) 01 and 04; Prescribing Information System (PIS), National Records of Scotland (NRS) Mortality and the Scottish Mental Survey 1947 (SMS1947). The datasets I used for dementia ascertainment were SMR 01 and SMR 04, PIS and NRS mortality. I outline these datasets, including a description of what data they cover in more detail in Table 5.7 below, alongside the data owner. As described in Section 3.6 (Data governance, access and permissions) and Table 3.1(Definitions of common governance terms use in routinely collected data), data ownership refers to the organisation maintaining the data in routinely

collected health data and not that they ‘own’ the data. Most of the routinely collected health datasets in Scotland have Public Health Scotland (PHS) as the data owner.

Dataset	Data owner	Began	Description
Scottish Morbidity Records 01 (SMR01)	PHS	1981	National hospital records. Episode level data. Covers general acute inpatient/day cases, including Geriatric Long stay inpatients for all acute specialities. Corresponds to the discharge from hospital (or discharge from one ward to another during the same hospital admission).
Scottish Morbidity Records 04 (SMR04)	PHS	1981	National hospital records. Episode level data. Covers mental health inpatients/day cases. All patients receiving care in the Scottish psychiatric hospitals at both admission and discharge.
Prescribing Information System (PIS)	PHS	1993	National community records. The national prescribing database. Includes prescribed and dispensed data. Individual patient level data.
National Records of Scotland Mortality (NRS Mortality)	NRS	1855	National community records. Statutory register of deaths of anyone in Scotland. Contains causes of deaths for anyone in Scotland. NRS was formally General Registrar Office (GRO). Multiple causes can be listed as ICD codes.

Table 5.7: Routinely collected health datasets in Scotland for dementia ascertainment in the *SMS1947 dementia data linkage study*

Dementia was ascertained from any records mentioning ICD-9 or 10 dementia codes (Table 5.9 and Table 5.10 in hospital discharge datasets (SMR 01 or 04), or from mortality datasets (NRS Mortality) if at any point a dementia ICD code was mentioned as a cause of death. For the PIS dataset, any mention

of one of the four anti-dementia drug prescriptions (Donepezil hydrochloride, Galantamine, Rivastigmine, Memantine hydrochloride) constituted a diagnosis of dementia (Section 5.5.1.3 below).

5.5.1.1 Scottish Morbidity Records (SMR) 01 and 04

The Scottish Morbidity Records (SMR) cover several national datasets. They are compiled and governed by Public Health Scotland (PHS), and previously by Information Services Division of NHS (ISD NHS) [683]. They include all information about patients collected from Scottish Hospitals. They consist of SMR 00 – SMR 11 datasets, although the Scottish Birth Record (SBR) replaced SMR11 (Neonatal Inpatient dataset) in 2002 [669]. Each dataset covers a specific area. Some are not relevant to my research e.g., SMR 02 covers Maternity inpatient and Day cases. For the purpose of my research, I used two commonly used datasets for the adult population to investigate dementia. These were the SMR01 and SMR04 datasets. Many of the SMR datasets are linked to research datasets, including population-based studies such as the Scottish Longitudinal Study (SLS) [685].

The SMR 01 and 04 datasets have conditions listed as main or other conditions. In the SMR 01 dataset there is one ‘main condition’ and up to five ‘other conditions’. In the SMR 01 dataset, ‘main condition’ refers to the condition listed on the discharge summary which is the principal cause (diagnosis) leading to the admission. However, in the SMR 04 dataset there are two variables for ‘main condition’ (‘main condition’ and ‘admission main condition’) followed by four further ‘other conditions.’ The ICD codes for these two main condition variables do not always match as a person may be admitted with one reason but following investigations and treatment the ‘main condition’ reason may have changed. Therefore, for the SMR 04 dataset we used the variable ‘main condition’ as this reflected the true diagnosis what was recorded on the discharge, and therefore was more comparable to the discharge ICD code in SM01 and more reflective of the ‘true’ diagnosis leading to admission.

As mentioned in Section 5.5.1, dementia was ascertained by any mention of ICD-9 or 10 dementia diagnostic codes (see Table 5.9 and Table 5.10) in hospital (SMR 01 or 04). In the SMR dataset, some entries had ICD 10 codes for dementia followed by a letter ‘A’ or ‘D’. These represent an asterisk (A) or dagger (D) and highlight another code linked to the first ICD code. To ensure complete extraction of all dementia related ICD codes, and none excluded due to the additional D or A, I wrote code in R to extract any ICD code for dementia on its own, or with an ‘A’ or ‘D’ following the ICD code.

5.5.1.2 National Records of Scotland (NRS) Mortality

The National Records of Scotland (NRS) produces information in Scotland on deaths [521]. They produce death information for Local Authority and Health Board areas. These are anonymised and publicly available. However, they do not publicly provide information at smaller geographical scales than Local Authorities. The NRS Mortality records available for my research within the National Safe Haven covered deaths between 4th June 1947 (the date the SMS1947 cohort was created) and 31st December 2019. As censorship of the SMS1947 was until 2018, SMS1947 and linked CHI data was only until 2018.

The death certificate in Scotland is called the Medical Certificate of the Cause of Death (MCCD) [521]. It is recorded as a Part 1 and Part 2 by the medical professional completing it. There is clear guidance for this in Scotland [521]. Depending on the judgement of the certifying doctor, some medical conditions of the deceased will be recorded and some not, depending on the relevance and contribution to the cause of death.

The NRS use the MCCD information to create a death registration record using ICD 10 codes for the conditions and diseases mentions in text on the MCCD. Anything recorded on the MCCD should be recorded on the death registration record. The only exception to this would be if there are more causes of death recorded on

the MCCD than positions to record ICD codes on the electronic record. All ICD codes are then listed within the dataset under numerous ‘cause_of_death_codes’ variables. There is space for a maximum nine codes for the underlying cause of death for each NRS Deaths record.

There can be up to nine ICD codes listed as cause of death in the NRS Mortality dataset. An ICD code for dementia can feature in any of those points. Theoretically, an ICD code for dementia could feature in more than one entry if someone has two sub-types of dementia that were coded separately.

Using R code I created code listing all the ICD 9 and ICD 10 codes for dementia from Table 5.9 and Table 5.10 above. I then ran this R code within the full NRS Mortality dataset within the database population. Once this analysis was completed, I created a file of dementia cases from death data. As described in the next Section 5.5.2.1 I subsequently created a ‘first record of dementia’ for this data file.

5.5.1.3 Prescribing Information System (PIS)

The Prescribing Information System (PIS) is an NHS dataset. It provides information on prescriptions from 1993 across Scotland. It includes data on prescribing and dispensing at individual (patient) levels. It also has financial information on the prescription/dispensing data. Items are described at the British National Formulary (BNF) item description level [472]. The data are released monthly. It includes prescriptions for General Practice data and also aggregated by NHS board for other prescriber location types [686].

Any anti-dementia drug prescriptions in the community prescribing data set (PIS) were used as a proxy for a diagnosis of dementia. These drugs are specific for dementia and can only be used clinically if someone has a diagnosis of dementia. Therefore they are a useful proxy for ascertaining a diagnosis of dementia which might not be recorded via ICD (versions 9 or 10) codes in the datasets I had

(secondary care) but in other datasets that were not accessible for my project. Therefore in the prescribing database (PIS) we extracted information on whether an individual was prescribed any of the four anti-dementia drugs listed in Table 5.8 below. R code was written to extract these specific names and cross check with British National Formulary (BNF) codes [472]. All the anti-dementia drugs were in Chapter 4.11 of the BNF and therefore all BNF codes relevant to the drugs started with the digits ‘0411...’[472].

At the time of my research, the drugs in the National Safe Haven for my study all had a corresponding BNF drug code [472]. I was able to extract the anti-dementia drugs through both the generic names (Donepezil hydrochloride, Galantamine, Rivastigmine, Memantine hydrochloride) and cross check that I had extracted all relevant entries by using the BNF codes of these drugs (as highlighted below in Table 5.8). I extracted by generic pharmaceutical (non-trade) drug name to ensure accurate extraction of all key dementia drugs for a diagnosis of dementia.

Drug	BNF code	Drug class
Donepezil hydrochloride	0411000D0	Acetylcholinesterase inhibitor
Galantamine	0411000F0	Acetylcholinesterase inhibitor
Rivastigmine	0411000E0	Acetylcholinesterase inhibitor
Memantine hydrochloride	0411000G0	Glutamate receptor antagonist

Table 5.8: Anti-dementia drugs used for dementia ascertainment in Scotland

ICD 10 codes for dementia were obtained from merging the published recognised list in datashare, the HDRUK phenotype library, and NHS data dictionary to ensure the most comprehensive list of dementia ICD codes [687, 688, 689, 690]. Table 5.9 and Table 5.10 show the ICD versions 9 and 10 codes used for ascertaining dementia.

ICD 9	Description
094.1	Dementia - syphilitic
290.0	Dementia - senility/old age
290.9	Dementia - idiopathic presenile or senile
290.10	Dementia - presenile simple type/uncomplicated
290.11	Dementia - presenile acute confusional state/delirium
290.12	Dementia - Binswanger's also presenile paranoid type
290.13	Dementia - presenile depressed type
290.20	Dementia - senile with delusional features
290.21	Dementia - senile with depressed type
290.3	Dementia - senile with acute confusional state/delirium
290.40	Dementia - vascular - arteriosclerotic (simple type)/multi-infarct
290.41	Dementia - arteriosclerotic with acute confusional state/delirium
290.42	Dementia, unspecified, with delusions
290.43	Dementia - arteriosclerotic with depressed mood
290.9	Dementia - degenerative
291.2	Dementia - alcohol-induced persisting
292.82	Dementia - drug-induced persisting
294.8	Dementia - dialysis
294.10	Dementia - frontotemporal without behavioural disturbance
294.11	Dementia - frontotemporal with behavioural disturbance
294.20	Dementia - unspecified
294.21	Dementia with behavioural disturbance
295.0	Dementia - simple type/simplex (acute) - see also schizophrenia
295.1	Dementia hebephrenic (acute)
295.2	Dementia - catatonic (acute)
295.3	Dementia - paranoid/paraphrenic
295.9	Dementia - developmental
319	Dementia - congenital (also in disability, intellectual)
330.1	Dementia - cerebral lipidoses with/without behavioural disturbance
331.0	Dementia - Alzheimer's with/without behavioural disturbance
331.19	Dementia - frontotemporal or frontal
331.82	Dementia - Lewy body with/without behavioural disturbance/dementia with Parkinsonism
345.9	Dementia - epilepsy with/without behavioural disturbance

Table 5.9: ICD 9 diagnostic codes used dementia ascertainment in Scotland

ICD 10	Description
A810	Sporadic Creutzfeldt Jakob disease
F051	Delirium superimposed on dementia
F00	Dementia in Alzheimer disease
F000	Dementia in Alzheimer disease with early onset
F001	Dementia in Alzheimer disease with late onset
F002	Dementia in Alzheimer disease, atypical or mixed type
F009	Dementia in Alzheimer disease, unspecified
F021	Dementia in Creutzfeldt-Jakob disease
F024	Dementia in human immunodeficiency virus (HIV) disease
F022	Dementia in Huntington disease
F02	Dementia in other diseases classified elsewhere
F028	Dementia in other specified diseases classified elsewhere
F023	Dementia in Parkinson disease
F020	Dementia in Pick disease
F013	Mixed cortical and sub-cortical vascular dementia
F011	Multi-infarct dementia
G308	Other Alzheimer disease
F018	Other Vascular dementia
F012	Sub-cortical vascular dementia
F03	Unspecified dementia
F03X	Unspecified dementia
F01	Vascular dementia
F010	Vascular dementia of acute onset
F019	Vascular dementia, unspecified
F106	Mental and behavioural disorders due to use of alcohol - amnesic syndrome
G30	Alzheimer disease
G300	Alzheimer disease with early onset
G301	Alzheimer disease with late onset
G308	Other Alzheimer dementia
G310	Circumscribed brain atrophy
G311	Senile degeneration of brain
G318	Other specified degenerative diseases of nervous system
I673	Binswanger's disease

Table 5.10: ICD 10 diagnostic codes used for dementia ascertainment in Scotland

5.5.2 Creating the study population

The SMS1947 ($n=70,805$, the *source* population) had previously been traced and CHI-linked to the routinely collected health data within Scotland's NSH ($n=53,037$, the *database* population) by data analysts at eDRIS as described in Section [5.3.3.1 Data linkage of Scottish Mental Survey 1947](#) as part of the larger Project Two, '*Mental health outcomes in SMS1947*'. This *database* population ($n=53,037$) was available in Scotland's National Safe Haven prior to the start of my study.

Next I created the *study* population from the above *database* population which also had geographical data available. Only those with a Urban-Rural 8-fold classification 2003-2004 ($n=16,910$) variable within the *database* population could be included in the *study* population, and analysis. Figure [5.4](#) highlights a flowchart of the process of creating the *study* population from the original *source* population.

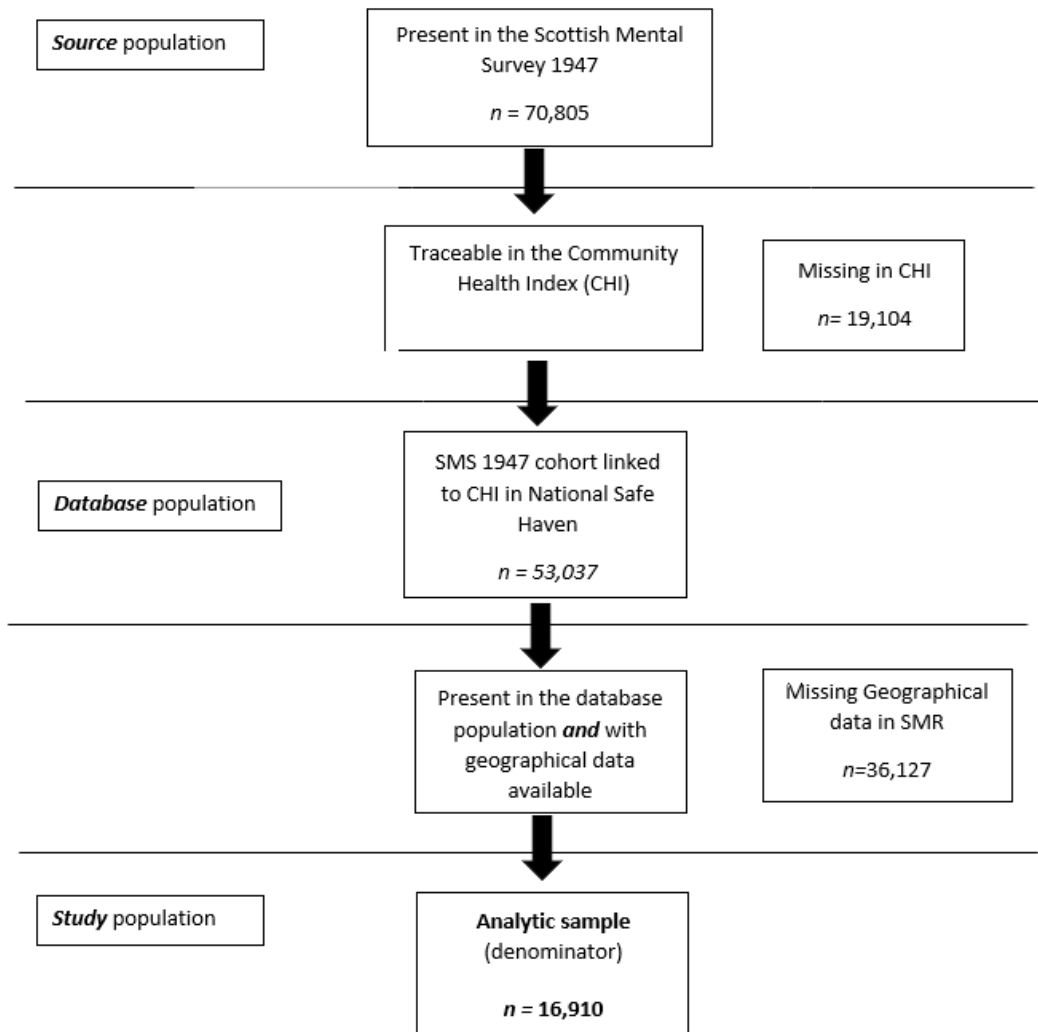


Figure 5.4: Flowchart of the process of creating the **study** population from the original **source** population

Coding using R software was used to undertake the creation of the **study** population. The **study** population is also referred to as the denominator population. I have described these layers of population hierarchy in Chapter 3 (Table 3.4 from Benchimol et al. [433]). and outline this in Table 5.11 in relation to my *SMS dementia data linkage study*.

Population type	Description	Size (n)
Source population	The historical Scottish Mental Survey 1947 (SMS1947) is the source population from which the database is derived. The population to make inferences from	70,805
Database population	The population derived from source population (SMS1947) who have been traced by the data linkage team and CHI-linked to routinely collected health data. The population available as a database in the National Safe Haven.	53,037
Study population	The population I created (using R coding) within the National Safe Haven to investigate my research question. Population consists of those from the database population whom also have geographical data present	16,910

Table 5.11: Population hierarchy in the *SMS dementia data linkage study*

My numerator population (people with dementia, PwD) were all those who were found to have dementia in any one of three datasets: ICD-10 codes in either 1) hospital discharges (SMR 01 or 04 datasets) or 2) death records (NRS death records dataset); or 3) prescription of an anti-dementia drug (PIS dataset) as discussed above in Section 5.5.1. All individuals were also flagged as alive or dead which allowed calculation of their age at the end of censoring in 2018 for those known to still be alive. An individual could have multiple records of dementia. I discuss this below in Section 5.5.2.1 including the methods for creating one entry for a diagnosis of dementia for each individual with their earliest record of dementia within any dataset. Individuals were flagged as 0 = no dementia (PwoD) and 1 = dementia (PwD).

5.5.2.1 Creating the first record of dementia

Once all people with dementia were flagged in the *database* population, each individual had multiple records of dementia linked with their unique identifier. The next step was to create a dataset where each individual with dementia was flagged once as having dementia and that their earliest record of dementia, and corresponding age at that time, was captured.

Of the $n=16,910$ individuals in the *study* population, the flagged data file was extremely large given these multiple records for each of the individuals with dementia. It was a long format file where an individual featured multiple times in multiple rows for each interaction with any of the datasets where dementia was recorded and they were flagged in those multiple datasets. This could include multiple flags within the same admission in the SMR datasets.

Multiple records can be created in SMR if an individual moves wards within a hospital during the same admission which generates a discharge summary from each ward and SMR entry for each ward move. Therefore, dementia could be flagged multiple times as discharges during the same admission in the SMR dataset as well as other points across a person's life time. Likewise, repeat prescriptions of an anti-dementia drug can result in multiple entries for the same individual within the PIS dataset for every prescription generated over the person's lifetime.

Multiple records must be reduced to one record per individual. This had to be done carefully so that the earliest time point that dementia features in any of the datasets was not lost when all the datasets were merged. To calculate the earliest time point that dementia features in any of these datasets the first step was to reduce the records in the individual records down to the earliest entry for each datasets for each individual with dementia.

The individual's corresponding age at that first mention of dementia was taken

as their age at diagnosis of dementia. This is in keeping with methodologies in other studies using routinely collected health data to ascertain dementia [152]. It is impossible to be certain that this *is* the first recording of dementia, but this assumption is commonly accepted methodologically, within the limitations of routinely collected health data largely using secondary care data, which I discuss in interpreting the results.

The only dataset where dementia would only be recorded at one point in time was the NRS mortality dataset. Each individual, if they had died and had dementia recorded, would only have one date for a dementia diagnosis for this dataset. This date would be their age at death.

For other datasets (SMR and PIS) careful de-duplication of records was required to ascertain the first mention of dementia, and age of diagnosis, and ensure that data was not lost when merging.

In Figure 5.5, I outline the methods of the first steps of 1) flagging people with dementia within each of the datasets if they had a diagnosis of dementia, and 2) identifying the first record of dementia recorded for each individual with dementia in each datasets separately, and finally 3) reducing the multiple records to one record per individual (keeping only the earliest date of dementia) in each of the datasets separately.

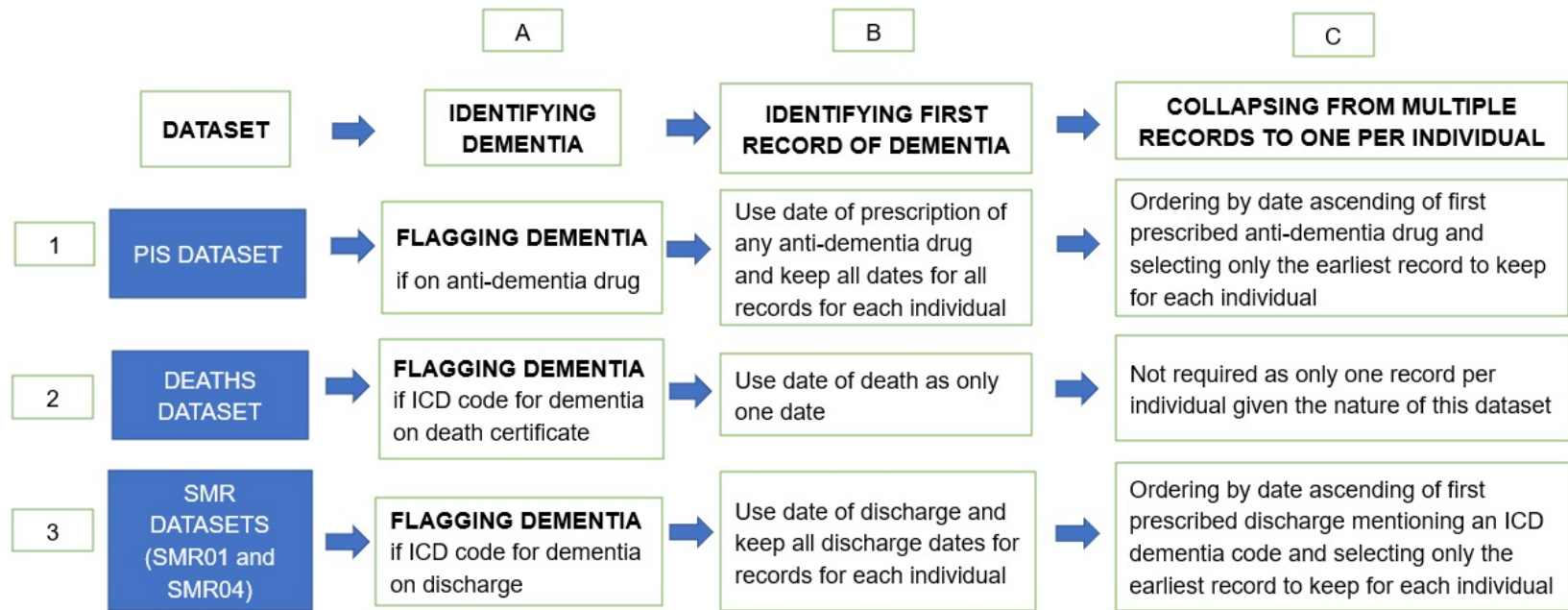


Figure 5.5: Flowchart depicting the process of dementia flagging in each dataset

Once the above process had taken place, I then merged the datasets together. Individuals could be in multiple datasets. The merged dataset needed to have each individual with dementia flagged once with the earliest data of diagnosis from any of the datasets. It was important to cross-check at multiple stages when creating R code to ensure individuals are not removed or ‘lost’ in the multiple steps to ‘collapse’ a long format dataframe (multiple entries in rows for each individual with multiple records for dementia) to a wide format dataframe (one individual, one data for dementia diagnosis) in R. More details on this process are outlined in [Appendix 5](#).

For the age of first dementia in the SMR datasets, I could have used the age at admission or the age at discharge within the SMR datasets. I selected the age at discharge as the SMR ICD-9 and ICD-10 codes refer to the diagnosis at discharge, and in case a diagnosis of dementia was given during that admission this was a more appropriate age to select.

5.5.3 Merging the flagged datasets

Using R code I proceeded with merging the flagged datasets in stages. As there were two SMR datasets (SMR01 and SMR02), these two datasets were first merged to form one SMR dataset flagged for dementia. Following this, the three dementia-flagged datasets (SMR, PIS and NRS mortality) were merged together.

Once all the datasets were merged, I used R code to ‘group’ them by individual ID. This resulted in each individual having all their multiple records one after another in multiple rows with their corresponding records. Using ‘filter’ R code, those records with a dementia code and flagged with dementia were selected. Finally, using ‘ascending order’ R code I ascended them in order with the earliest date of dementia (‘first_age’ column) in ascending order. This allowed me to select the first row for each individual who was flagged with dementia and had their earliest record of dementia as the first row with their individual ID.

Table 5.12 illustrates this using a ‘dummy’ ID and ‘first_age’ as an example as the actual data cannot be outputted from the National Safe Haven due to confidentiality. In the table below, the merged dataset is grouped by ‘Individual ID’ and ‘filtered’ to all those with dementia (1) in the dementia_flag. The ‘first_age’ for each record is put in ‘ascending order.’ Only the first record (with the youngest age of dementia recorded) is selected from the merged dataset for each individual. In Table 5.12, I have highlighted in yellow the ones that would be kept. In addition, I also created a column labelled ‘source’ to distinguish which dataset the entry came from when the three cleaned and flagged datasets (PIS, Deaths for NRS mortality, SMR) were merged together.

Individual ID	Dementia flag	First age (months)	Source
000001	1	899	SMR
000001	1	900	PIS
000001	1	901	PIS
000001	1	903	Deaths
000002	1	700	SMR
000002	1	703	SMR
000002	1	704	PIS
000003	1	800	PIS
000003	1	801	PIS
000003	1	802	PIS
000003	1	809	SMR
000004	1	888	SMR
000004	1	899	SMR

Table 5.12: Mock example of grouping, filtering 4 individuals with dementia from records based on first recorded data of dementia within any of the datasets

Those selected were saved as the ‘dementia cases’ dataset (e.g. all those individuals flagged with dementia, PwD, and only one record of dementia per individual which was their earliest point in the datasets with a dementia diagnosis). This

‘dementia cases’ dataset was for the whole *database* population. The final step was to merge this with the *study* population (those who also had geographical data available for the analysis of my research question). This meant some individuals in the ‘dementia cases’ dataset (within the *database* population) were intentionally removed when merging with the *study* population. I outline this final step in the process in Figure 5.6.

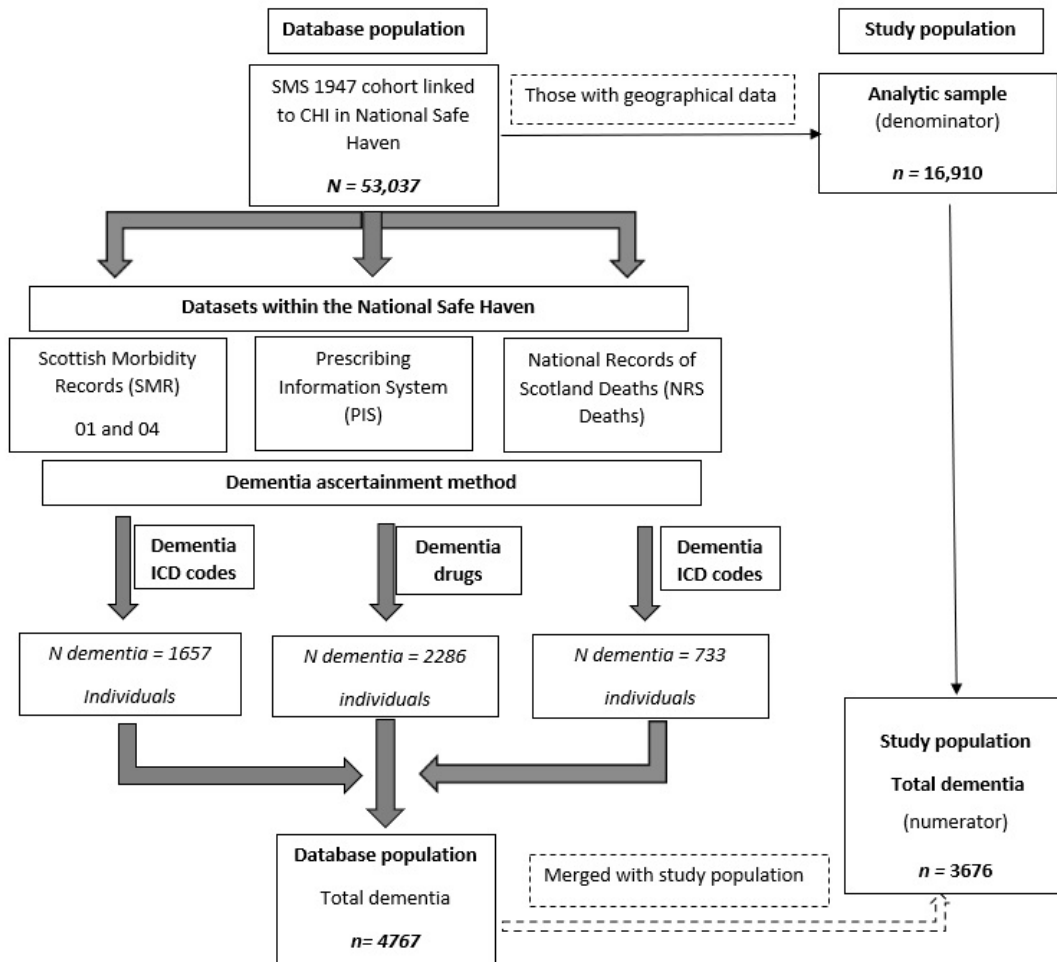


Figure 5.6: Flowchart of process ascertaining dementia from the *database* population and merging with the *study* population

5.5.4 Data cleaning and missing data

Using R code, data was first cleaned to remove white spaces. White spaces are gaps, spaces, tabs or other white-space characters from a string of data. Keeping

white space can lead to incorrectly processing it when coding.

Data files can often have errors in data entered. Therefore, we cleaned the data to remove individuals over a certain age as explained here. Censorship (the point at which final data linkage of the SMS1947 with routinely collected health data) was 2018. Participants were aged eleven in 1947, with most born in 1936 but a few born either side (1935-1937).

Individuals over a certain age (1044 months, 87 years) were removed, as they should not be in the *database* population. The oldest age at the time of cognitive assessment was 137 months (11.4 years old). There were 907 months (75.6 years) between the intelligence test at age eleven (137 months, or 11.4 years, old at oldest age), and linking to prescription data. I accepted that the oldest could be a maximum of 1,044 months (87 years) old. This was in line with past published research using this SMS1947 CHI-linked data in the NSH [691, 692]. All prescriptions for anyone aged over 1044 months (87 years old), as well as those with NA (no entry), were removed. Individuals outside the expected age were likely due to clerical error on data entry, which is a known risk in all datasets.

‘Dummy’ prescriptions (prescriptions created by pharmacy for training purposes) were removed. Lastly, some prescriptions had the wrong truncated BNF codes. I excluded these rather than impute missing data by giving them a generic code from similar medications as this might introduce bias. We did not impute missing data as we could not be sure the data was missing at random and did not want to introduce bias.

5.5.5 Statistical analysis

I extracted counts of all-cause dementia and within each dataset (de-duplicated so individuals not records were extracted) for their first recording of dementia in the datasets within the *database* population. An individual was classed as having dementia irrespective of their death status at the end of censorship (2018)

- e.g. an individual was classed as having dementia if they had died before 2018 and had dementia recorded in the SMR, PIS or NRS mortality datasets. In the analysis I refer to this as ‘diagnosis of dementia (any death status).’

Using those people flagged (coded as 1) with dementia within the *database* population using all four datasets (SMR 01, SMR 04, PIS and mortality). Following this, I calculated the prevalence of all-cause dementia (numerator population) within the *study* population by combining these datasets and linking them back to the *study* population (SMS1947 participants whom were CHI-linked with geographical data).

Within the *study* population (all with geographical data within the SMS 1947 who were CHI-linked to routinely collected health data), I extracted descriptive data for those with dementia (PwD, cases), and those without dementia (PwoD, controls) including:

- Sex
- Urban-Rural 8-fold classification at first SMR admission
- SIMD 2004 decile at first SMR admission (measure of deprivation)
- Dementia – yes/no
- Moray House Test score (intelligence)

I compared the basic characteristics of the participants with and without dementia. Statistical analysis was undertaken using R version 4.2.2 within the secure NSH. To be outputted, R code must be pre-agreed by eDRIS as suitable to be output from the NSH. I have not included this in the Appendix, but it is available on request.

Cox proportional hazards (CPH) models were undertaken on the study population. The outcome of interest (status) for the CPH models was dementia/no dementia (any death status).

The time used in the models was the first age of dementia/no dementia. For those *without* dementia the time (age) was the age at the end of censorship (if still alive at the end of censorship in 2018) or age at death if died before the end of censorship (and no dementia recorded on their death certificate). Urban was used as the reference point. A comparison on the outcome of dementia (any death status) in relation to age and varying between urban or rural living at the first mention in the SMR records was modelled. This was then adjusted for the co-variables of interest: sex, intelligence, SIMD (deprivation).

A Cox model of proportional hazards regression analysis was calculated to examine time to event (dementia, any death status) using these two predictors:

1. Rural/urban location (using the 8-fold Urban-Rural 2003-2004 classification system[\[663\]](#))
2. Age in months (details below)

A Cox Proportional Hazards (CPH) model was used over logistic regression as there was broader rural/urban information across time. It was felt that a CPH model would be a better statistical model to investigate the impact of rural/urban living, time and outcome/status (dementia – any death status). Whilst age was included in months, it was less important as this was a group (SMS1947) with largely similar age due to the nature of the historical survey.

For those with dementia (PwD, cases), the age in months was taken as the age when dementia was first recorded in any of the three datasets (SMR, PIS or NRS mortality). For those without dementia (PwoD, controls), age was taken as: 1) the age in months at the time of death (if died) or 2) the age in months at the end of censoring (2018) if alive. Age was later converted from months to years.

Prior to running the CPH regression model in R, I assessed the model's assump-

tions, including the proportional hazards assumption. The model in R [`residuals.coxph()`] assumes that the baseline hazard function is constant over time and that the hazard ratios for the predictor variables are constant over time (proportional hazards assumption). The R function [`residuals.coxph()`] is based on these assumptions. If the proportional hazards assumption were violated the residuals may not provide accurate assessments of model fit, or may not be possible to interpret in the usual manner, and additional strategies such as stratification would need to be implemented.

In particular, sex is known to influence dementia risk (female sex has an increased risk of dementia). Therefore, I checked the sex and Urban-Rural 2003-2004 classifications for differences on dementia (outcome).

I undertook unadjusted and adjusted models of CPH. I adjusted for several covariates: sex, levels of deprivation (SIMD 2004 deciles), age eleven intelligence score (Moray House test age eleven standardised) and Urban-Rural classification at different levels of rural/urban as discussed below.

I kept these as the eight separate classifications but also grouped them as 1) Urban (1-5) versus Rural (6-8), 2) Accessible (1-3 and 6) versus Remote (4-5, 7-8), and 3) All Urban (1-5) classifications versus the three different Rural classifications individually 6, 7 and 8. All outputs from the NSH were subject to review and output checks by the eDRIS research coordinator to ensure anonymity, and that they remained non identifiable, prior to approval. Any low values were changed to 'less than' a higher number to ensure no small numbers were outputted in line with their regulations.

5.6 Ethics

Ethical approval had already been discussed with the appropriate authorities by the Principal Investigator prior to my involvement in using Project 2 '*Mental*

health outcomes in SMS1947' for my research question. My study was aligned with the aims and objectives of the overall project.

Berkshire NHS Research Ethics Committee waived the need for full ethical review of this work (application 18/SC/0505). This work is instead covered under the National Safe Haven Generic Ethics approval, run by Public Health Scotland and approved by the East of Scotland NHS Research Ethics Service. The HSC-Public Privacy Benefit Panel (HSC-PPBP) approved the study to take place within the National Safe Haven.

5.7 Chapter conclusions

This chapter has given an overview of the location (Scotland), the scales (Scottish geographical and deprivation scales) and the historical survey (Scottish Mental Survey 1947, SMS1947) that I have used in my research (the *SMS1947 dementia data linkage study*), the results of which can be found in the next thesis [Chapter 6: Results of the SMS1947 dementia data linkage study](#).

In summary, I have outlined the methodology for this research including: how dementia was ascertained, the stages in creating the study population, how the earliest entry for a dementia diagnosis was selected, the statistical analysis used, and highlighting the access and permissions process in the planning of this study.

Chapter 6:

Results of the SMS1947 dementia data linkage study

6.1 Chapter introduction

In this chapter, I will present the results of my research investigating dementia outcome with rural or urban living in the the Scottish Mental Survey 1947 (SMS1947) which was CHI-linked to routinely collected health data. This was called the *SMS1947 dementia data linkage study*. The methodology for this has been outlined in the prior chapter (Chapter 5). This Chapter will now focus on these results with discussion of the findings. To begin, I revisit the research aims and objectives to outline the focus of the research.

6.2 Research aims and objectives

The aim was to establish who developed dementia and the variation between rural and urban areas of Scotland, using data from the SMS1947 historical survey linked to routinely collected health data in Scotland within the National Safe Haven (NSH). The research question was: *“Is there a difference in dementia rates between rural and urban areas of Scotland in the historical SMS1947 CHI-linked to routinely collected health data.”*

As highlighted in the Methods Chapter for this research (Chapter 5), my hypothesis was that dementia would be higher in rural areas versus urban areas in line with my systematic review in Chapter 2 and in keeping with a similar increase in the North (predominately more rural) versus South (predominately more urban) Scotland in previous research undertaken on the the older Scottish Mental Survey 1932 [350].

My objectives, as discussed in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), were:

1. From the **database** population (participants in the SMS1947 who have CHI-linked routinely collected healthcare data) create a **study** population consisting of those in the **database** population who also have geographical data available in Scotland's NSH.
2. Ascertain dementia within a) the **database** population (SMS1947 CHI-linked to routinely collected health datasets); and b) the **study** population (a sub-set of the SMS1947 CHI linked to routinely collected health datasets who also have geographical data available)
3. Investigate whether rural or urban location is linked with increased or decreased dementia rates (dementia diagnoses) in the **study** population

I will refer to this historical survey data (SMS1947) linked to routinely collected data as the '*SMS1947 dementia data linkage study*' throughout the rest of this Chapter and thesis.

6.3 Results

For this chapter I will refer to those with dementia as 'person with dementia' (PwD) and those without as 'person without dementia' (PwoD).

6.3.1 Denominator population

The denominator population (**study** population) was created from *first* the **source** population and then *secondly* the **database** population. This section outlines the numbers in each of the populations that are created in different steps of the data linkage. I created the **study** population (denominator) using R code within the National Safe Haven (NSH). This R code must undergo output reviews before being released from the NSH so is available on request.

Of the SMS 1947 historical survey (known as the *source population*), $n=70,805$, $n=53,037$ were traced to health records (known as the *database population*), and were available in Scotland's NSH. From this *database* population I excluded those who were missing geographical data, and only included those who specifically had the Urban-Rural Classification 2003-2004 variable. This was my *study population* also referred to as the *denominator* within data science, ($n=16,910$) as outlined in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#) Table 5.11.

6.3.2 Numerator population

Having created the *study* population (denominator), the next stage was to create the numerator (PwD) within this population. I have outlined the methods of how I created my study population and ascertained dementia in detail in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), Section 5.5.2 and Figure 5.6. To summarise, this required using the *database* population to flag all with dementia and then merging with my smaller *study* population (those that had geographical data available). This approach also allowed me to consider the number of people with dementia that were excluded from the *study* population as they were missing geographical data. As discussed in Chapter 5 (Section 5.5.2.1, to avoid counting an individual multiple times I narrowed the dementia ascertainment to their first record of dementia in any of the datasets.

In Table 6.1 I report the results of whom I ascertained dementia on ($n=$) and which dataset the individuals first had dementia recorded. As dementia was ascertained within the *database* population and then merged with the *study* population I report both dementia numbers in both populations. Some individuals featured in more than one dataset as would be anticipated (e.g., those with a dementia diagnosis on hospital discharge may also have prescription of an anti-dementia drug, or if they had died, have dementia recorded on their death certificate). Within the *study* population, the majority of dementia diagnoses were first recorded in PIS (notably more for females $n=908$, than males $n=567$)

and SMR 01, though all sources contributed unique cases.

Dataset where dementia first recorded	Database population ($n=53,037$)	Study population ($n=16,910$)
	Dementia, PwD $n= 4,767$ (8.9%)	Dementia, PwD $n= 3,676$ (21.7%)
SMR 01		
Male	606	601
Female	619	612
SMR 04		
Male	228	226
Female	204	202
PIS		
Male	903	567
Female	1,383	908
NRS deaths		
Male	302	234
Female	431	326

Table 6.1: Datasets where dementia first recorded for the *study* and *database* populations in the *SMS1947 dementia data linkage study*

6.3.3 Study population

The *study* population consisted of $n=16,910$ individuals (CHI-linked SMS1947 individuals with geographical data). Of theses, $n=61,118$ individuals (36.2%) were alive at the end of censoring (year=2018), and $n=10,792$ individuals had died.

The *study* population was split across all geographical classifications of the Urban-Rural 8-fold 2003-2004 Classification System from classification 1 (large urban areas) to classification 8 (very remote rural areas). Most of the *study* population lived in large urban areas ($n= 7,149$, 42.3%). Distribution across all 1-8 Urban-rural classification for the *study* population is shown in Table 6.2.

Class	Class Name	Description	Study population (<i>n</i>) (% total study pop.)
1	Large Urban Areas	Settlements of 125,000 people and over.	7,149 (42.3)
2	Other Urban Areas	Settlements of 10,000-124,999 people.	4,957 (29.3)
3	Accessible Small Towns	Settlements of 3,000-9,999 people, and within a 30-minutes drive time of a settlement of 10,000 or more.	1,694 (10.0)
4	Remote Small Towns	Settlements of 3,000-9,999 people, and a drive time of >30 minutes but less than or equal to 60 minutes to a settlement of 10,000 or more.	306 (1.8)
5	Very Remote Small Towns	Settlements of 3,000-9,999 people, and with a drive time of >60 minutes to a settlement of 10,000 or more.	241 (1.4)
6	Accessible Rural Areas	Areas with a population of <3,000 people, and within a drive time of 30-minutes to a settlement of 10,000 or more.	1,774 (10.5)
7	Remote Rural Areas	Areas with a population of <3,000 people, and with a drive time of >30-minutes, but less than or equal to 60-minutes to a settlement of 10,000 or more.	366 (2.2)
8	Very Remote Rural Areas	Areas with a population <3,000 people, and with a drive time >60 minutes to a settlement of 10,000 or more.	423 (2.5)

Table 6.2: *SMS1947 dementia data linkage* **study** population (*n*) within each Urban-Rural classification

Adapted from Scottish Government Urban-Rural Classification 8-fold [663] information on class name and description to include my study population results.

People with dementia within the *study* population

Table 6.3 outlines the descriptive statistics of the *study* population and differences (male/female, urban/rural or intelligence score) between those with (PwD) and without (PwoD) dementia.

	Dementia, PwD <i>n</i> (%)	No Dementia, PwoD <i>n</i> (%)
Individuals	3,676 (21.7)	13,234 (78.3)
Sex (female)	2,048 (55.7)	6,907 (52.2)
Moray House Test Score (Intelligence score)	-0.15 (1.02)	-0.12 (0.99)
Urban-Rural 8-fold 2003-2004		
Urban categories	3,151 (85.7)	11,196 (84.6)
Rural categories	525 (14.3)	2,038 (15.4)

Table 6.3: Descriptive statistics of the *study* population with and without dementia

Moray House Test (MHT) score is an intelligence test which is standardised with a mean of 0 and standard deviation of 1 for the whole SMS1947 participants. Age is not included in this table as the cohort were all born within the same year (1947) and age (first age of dementia) was used as the variable for time in the Cox Proportional Hazards ratios below.

As all in the *study* population were born in the same year (1947), age was not reported between PwD and PwoD as age was used as a variable for time in the Cox Proportional Hazard ratios. Age within the PwD group was reviewed in more detail in relation to the age of *first recording* of dementia.

The average age of dementia diagnosis (first recording) in the *study* population was 76.9 years (median= 78.1 years). This was not significantly different from the

database population, mean= 77.3 years (median = 78.4 years). The minimum age of dementia first diagnosed in the **study** population was 44.4 years and the maximum was 83.9 years.

There was a drop in dementia ascertainment in the **study** population (PwD $n=3676$, 21.7%) from the **database** population (PwD $n=4,767$, 9.0%), due to the smaller **study** population size (those with geographical data present). This was a decrease in 1,091 people (22.9% reduction) with dementia not ascertained in the **study** population but found in the **database** population.

Undertaking a Pearson's Chi-squared for this difference showed it was significant: $X^2=1962.8$, $df = 1$, $pvalue < 0.05$. In the **study** population there were $n=2,048$ (55.7%) PwD who were female, and $n=6,907$ (52.2%) PwD who were male. This significant drop in ascertainment rate of dementia from the whole **database** population to the **study** population will be explored in the discussion Section 6.4.

In summary, 21.7% of the **study** population were diagnosed with dementia, and 8.9% of the **database** population were diagnosed with dementia, suggesting higher prevalence of dementia within the **study** population group. This will be discussed further in Section 6.4 Discussion. Despite a smaller **study** population size, the people with dementia that I was unable to ascertain with dementia in this population (compared with the **database** population) as a result of being excluded due to missing geographical data was <23%.

Rural/urban location amongst PwD and PwoD: study population

More of the **study** population lived in urban areas (Classes 1-5), $n=14,347$ (84.8%) than lived in rural areas (Classes 6-8), $n=2,563$ (15.2%). Of those with dementia (PwD) in the **study** population ($n=3676$, 21.7%), $n=3151$ (85.7%) PwD lived in urban areas (Classes 1-5), and $n=525$ (14.3%) PwD lived in rural areas (Classes 6-8).

The proportions (individuals living in rural and urban areas) were similar in the *study* population for individuals with dementia. If considering dementia as a reflection of the population size lived in then Urban_dementia percentages (Urban_dementia $n=3,151$, total urban population $n=14,347$, % of dementia present in urban areas representative to the urban population size = 22.0%) were proportional to the Rural_dementia percentages (Rural_dementia $n=525$, total rural population $n=2,563$, % of dementia in rural areas representative to the rural population size=20.5%).

Figure 6.1 presents the numbers of individuals within the *study* population (both those with, PwD, and those without, PwoD, dementia) within the 1-8 classifications of the Urban-Rural 8-fold 2003-2004 classification system. The distribution of PwD varied across the 8 classifications of Urban-Rural. Classifications 1 (large urban areas), 2 (Other urban areas), 3 (Accessible small towns) and 4 (Remote small towns) had most inhabitants (both PwD and PwoD) of the *study* population.

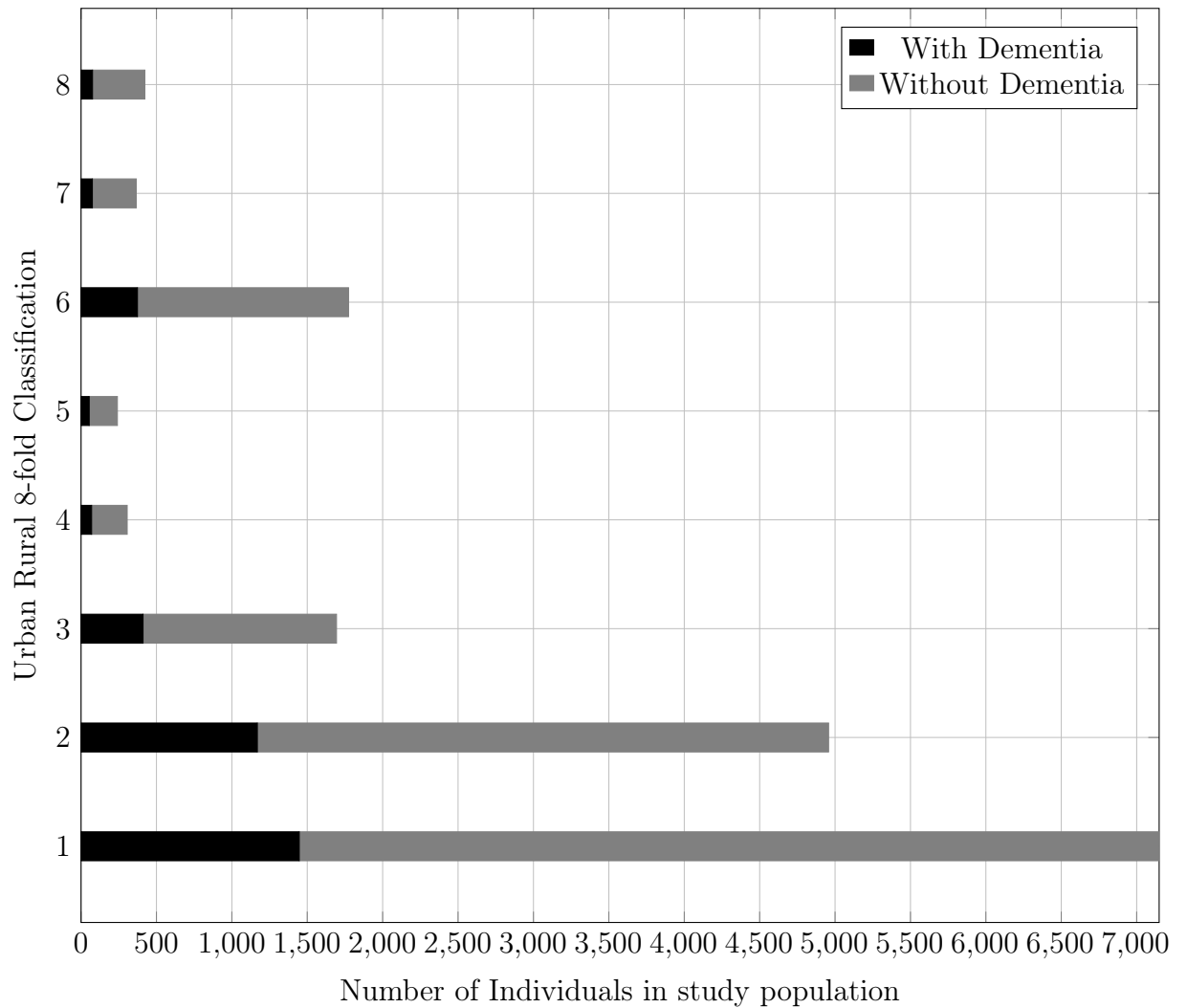


Figure 6.1: Study population: Individuals with and without Dementia by Urban-Rural 2003-2004 Classifications 1-8

Urban-Rural 8-fold 2003-2004 Classification system: 1 (Large Urban Areas - Settlements of 125,000 people and over); 2(Other Urban Areas - Settlements of 10,000-124,999 people); 3 (Accessible Small Towns - Settlements of 3,000-9,999 people, and within a 30-minutes drive time of a settlement of 10,000 or more); 4 (Remote Small Towns - Settlements of 3,000-9,999 people, and a drive time of >30 minutes but less than or equal to 60 minutes to a settlement of 10,000 or more); 5 (Very Remote Small Towns - Settlements of 3,000-9,999 people, and with a drive time of >60 minutes to a settlement of 10,000 or more); 6 (Accessible Rural Areas - Areas with a population of <3,000 people, and within a drive time of 30-minutes to a settlement of 10,000 or more); 7 (Remote Rural Areas -Areas with a population of <3,000 people, and with a drive time of >30-minutes, but less than or equal to 60-minutes to a settlement of 10,000 or more); 8 (Very Remote Rural Areas - Areas with a population <3,000 people, and with a drive time of over 60 minutes to a settlement of 10,000 or more).

Deprivation amongst PwD and PwoD in the *study* population

Figure 6.2 the numbers of individuals within the study population (both those with, PwD, and those without, PwoD, dementia) within the 10 deciles of the SIMD 2004.

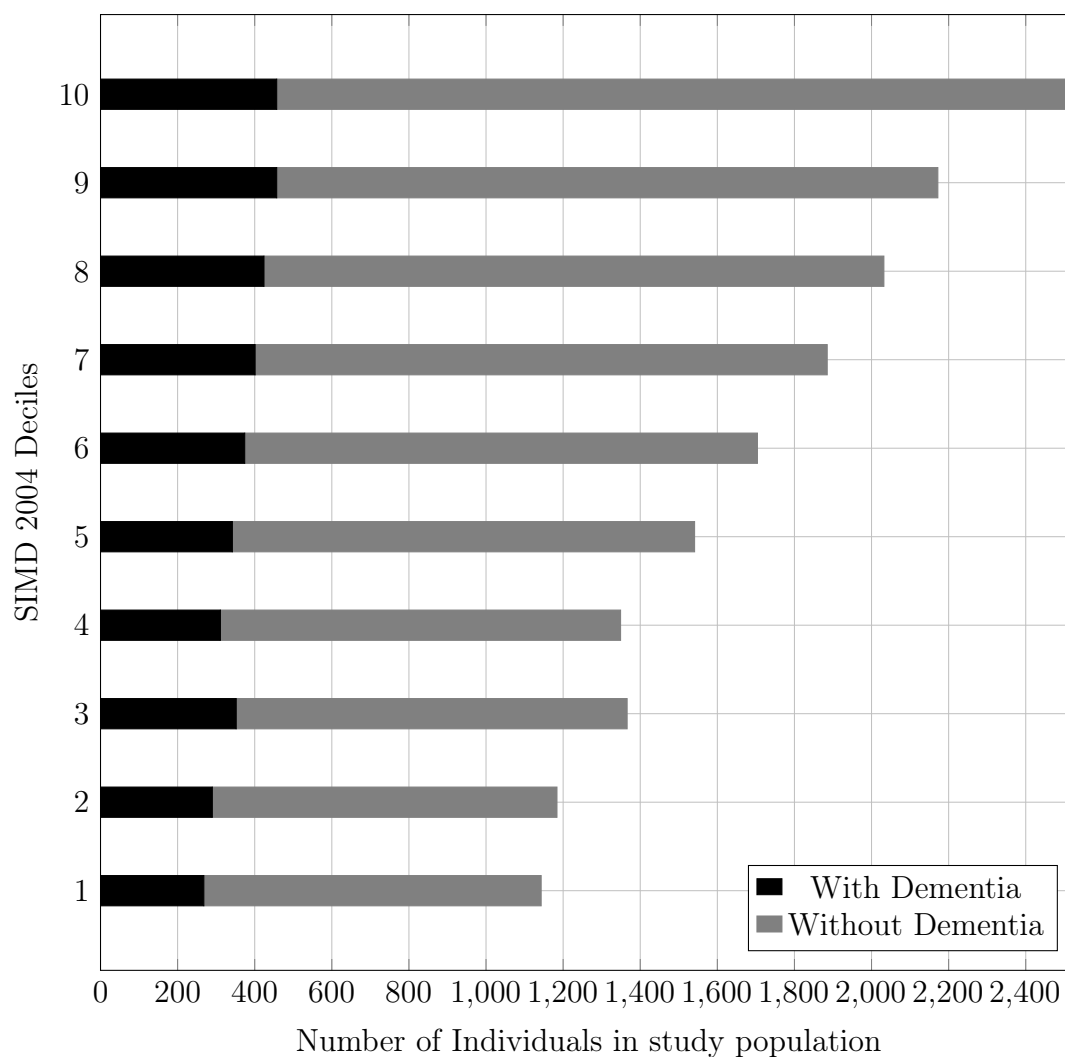


Figure 6.2: Study population: Individuals with and without dementia by SIMD 2004 Deciles

Deprivation is reported by SIMD deciles. SIMD decile 1 contains the most deprived 10% of data zones, and decile 10 contains the least deprived 10% of data zones.

6.3.4 Rural/urban living and dementia

As outlined in the methods for this research (Chapter 5, Section 5.5.5), Cox Proportional Hazard (CPH) Models were undertaken to examine the effect of rural versus urban living on a dementia diagnosis. The proportional hazards assumptions were not violated. Sex (female) did interact with a dementia diagnosis, but there was no interaction between sex and urban/rural living (binarized) with dementia. As there was no difference between sex and rural/urban effect on dementia, I did not stratify by sex as the CPH assumption was met.

Dementia diagnoses were higher in urban areas versus rural areas. Rural living (Classes 6-8) was associated with a 16% (CI 8-23%) lower rate (lower hazard ratio) of dementia than urban living. This does not change when adjusted for age, sex, intelligence or SIMD. Two CPH models are presented below in Tables 6.4 and 6.5. The remaining models are in 7.7 of this chapter. The models were adjusted for sex, sex and intelligence, sex and SIMD, and all three of the covariates in the final model, Model 5. Table 6.4 illustrates the Cox Proportional Hazard ratios for urban versus rural living as a binarized analysis whereby *all* urban (1-5) classifications are compared to *all* rural (6-8) classifications.

	Urban-Rural classification (levels)	Size (<i>n</i>)	Hazard Ratio	Lower 95% CI	Upper 95% CI
Model 1 (unad- justed)	Urban (1-5)	14,347	1 (Ref.)		
	Rural (6-8)	2,563	0.84	0.77	0.92
Model 2 (sex)	Urban (1-5)	14,347	1 (Ref.)		
	Rural (6-8)	2,563	0.84	0.76	0.92
Model 3 (sex, Intelligence)	Urban (1-5)	14,347	1 (Ref.)		
	Rural (6-8)	2,563	0.84	0.76	0.92
Model 4 (sex, SIMD)	Urban (1-5)	14,347	1 (Ref.)		
	Rural (6-8)	2,563	0.85	0.78	0.94
Model 5 (sex, Intelligence, SIMD)	Urban (1-5)	14,347	1 (Ref.)		
	Rural (6-8)	2,563	0.84	0.77	0.93

Table 6.4: Cox Proportional Hazard Models (binarized Urban vs. Rural classifications)

Urban is the reference. Urban classifications 1-5 (1=large urban areas; 2=other urban areas; 3=accessible small towns; 4=remote small towns; 5=very remote small towns), and rural classifications 6-8 (6=accessible rural areas; 7=remote rural areas; 8=very remote rural areas). Intelligence refers to the corrected and standardised Moray House Test, MHT, (standardised to a mean =0, SD = 1, for all SMS1947 participants), SIMD refers to the Scottish Index of Multiple Deprivation variable in deciles. Model 1 unadjusted; Model 2 adjusted for sex; Model 3 adjusted for sex and intelligence; Model 4 adjusted for sex and SIMD; Model 5 adjusted for sex, intelligence and SIMD.

To explore this variation between dementia in urban and rural areas, I considered whether the level of accessibility influenced the result in rural areas. Rural areas

classified as 6 are ‘accessible rural areas’ whereas the other two classes of rural areas (7 and 8) are ‘Remote Rural areas’ and ‘Very Remote Rural Areas’ respectively. I undertook sensitivity analysis by exploring the three rural areas (Classes 6,7,8) to understand if there were differences within the areas classed as rural. I compared the rural areas with all urban areas grouped together as one.

Table 6.5 illustrates the results of this Cox Proportional Hazard Model analysis which illustrates that whilst the results did not change significantly, the Hazard Ratio was lower for the more remote rural areas (i.e. less likely to have a diagnosis of dementia if living in the more remote rural areas). The unadjusted model (model 1) showed that the Hazard ratios were: 0.87 for ‘Rural-accessible’ (Level 6) areas (95%CI 0.78-0.96), 0.81 for ‘Rural-remote’ (Level 7) areas (95%CI 0.64-1.02) and 0.77 (95%CI 0.61-0.96) for ‘Rural-very-remote’ (Level 8) areas. This does not change significantly when adjusting for sex, intelligence and SIMD.

Overall, there was a clear direction in all models (unadjusted and adjusted) that showed the more rural and remote the region then the lower the Hazard Ratio was for a dementia diagnosis.

	Urban-Rural classification (levels)	Size (<i>n</i>)	Hazard Ratio	Lower 95% CI	Upper 95% CI
Model 1 (unadjusted)	Urban (1-5)	14,347	1 (Ref.)		
	Rural-accessible (6)	1,774	0.87	0.78	0.96
	Rural-remote (7)	366	0.81	0.64	1.02
	Rural-very-remote (8)	423	0.77	0.61	0.96
Model 2 (sex)	Urban (1-5)	14,347	1 (Ref.)		
	Rural-accessible (6)	1,774	0.86	0.77	0.96
	Rural-remote (7)	366	0.81	0.64	1.02
	Rural-very-remote (8)	423	0.76	0.61	0.96
Model 3 (sex, intelligence)	Urban (1-5)	14,347	1 (Ref.)		
	Rural-accessible (6)	1,774	0.85	0.76	0.95
	Rural-remote (7)	366	0.83	0.66	1.05
	Rural-very-remote (8)	423	0.77	0.61	0.98
Model 4 (sex, SIMD)	Urban (1-5)	14,347	1 (Ref.)		
	Rural-accessible (6)	1,774	0.88	0.79	0.98
	Rural-remote (7)	366	0.82	0.66	1.04
	Rural-very-remote (8)	423	0.77	0.61	0.96
Model 5 (sex, intelligence, SIMD)	Urban (1-5)	14,347	1 (Ref.)		
	Rural-accessible (6)	1,774	0.86	0.77	0.96
	Rural-remote (7)	366	0.84	0.67	1.06
	Rural-very-remote (8)	423	0.78	0.61	0.98

Table 6.5: Cox Proportional Hazard Models comparing all urban grouped together, against the individual rural classifications

Urban is the reference. Urban classifications 1-5 (1=large urban areas; 2=other urban areas; 3=accessible small towns; 4=remote small towns; 5=very remote small towns), and rural classifications 6-8 (6=accessible rural areas; 7=remote rural areas; 8=very remote rural areas). Intelligence refers to the corrected and standardised Moray House Test, MHT, (standardised to a mean= 0, SD= 1, for all SMS1947 participants), SIMD refers to the Scottish Index of Multiple Deprivation variable in deciles. Model 1 unadjusted; Model 2 adjusted for sex; Model 3 adjusted for sex and intelligence; Model 4 adjusted for sex and SIMD; Model 5 adjusted for sex, intelligence and SIMD.

In summary, my results showed that those born in 1936 in Scotland, and later living in a rural area, were linked with a lower hazard ratio (unadjusted HR 0.84, 95% CI 0.77-0.92 and adjusted for sex, intelligence and SIMD, HR 0.84, 95% CI 0.77-0.93) of developing dementia than if living in an urban area. Dementia was lowest if living within the most rural/remote region (unadjusted HR 0.77, 95% CI 0.61-0.96, and adjusted for sex, intelligence and SIMD, HR 0.77, 95% CI 0.61-0.98).

6.4 Discussion

I will now discuss my results including the strengths, limitations and challenges of the research. Additionally, I discuss future work planned to expand this research and consider possible hypotheses for the variation that could be investigated.

I found that, contrary to my hypothesis, dementia diagnoses were lower in rural areas, and higher in urban areas. This could represent:

1. the rural environment is protective for dementia,
2. the urban environment has risk factors for dementia,
3. diagnostic variation in dementia diagnoses between rural and urban areas in Scotland,
4. differences in the *study* population or geographical data.

I discuss these aspects in more detail in this section. I begin this discussion section by comparing my results with existing literature, followed by discussion of the reasons for variation and finally discussing the strengths, limitations and challenges of the research.

6.4.1 Comparison with existing literature

Dementia rates vary geographically as outlined in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#). My results in this chapter are contrary to my hypothesis that dementia rates would be higher in rural areas, derived from research previously investigating the older Scottish Mental Survey (SMS) 1932 (born in 1921 and sat the MHT, intelligence test, aged eleven in 1932 across Scotland) [350]. That research demonstrated geographical variation of dementia with higher odds of dementia in the north versus south of Scotland (the north is overall more rural than the southern/central belt of Scotland) [350]. They found no variation in dementia odds in childhood but substantial variation (increased odds in the North, more rural, versus South, more urban, of Scotland) for dementia odds by adulthood. This study had access to the county of school attended and residential location on first mid-life admission to hospital (by at least age 60 years) for the locations to model in their analysis. Those not admitted to hospital had location at death recorded as their location.

Additionally, a systematic review in 2012 showed dementia was higher in rural versus urban areas, as highlighted in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#) [340]. Further studies have shown increased dementia in rural areas, but with important considerations including socio-economic deprivation that might be influencing these results. Other studies have suggested no difference or even conversely increased dementia in urban areas.

One study in the USA investigated rural residence in childhood amongst African-Americans [235]. It found rural childhood residence was associated with an increased risk of Alzheimer's disease [235]. Importantly this was rural residence combined with the number of years of education, and the authors recognised that low education might be a marker for other influences in childhood such as socioeconomic or environmental influences [235]. This study also found that ru-

ral residence was highly correlated with low education levels (of the individuals reporting rural residence $n=79$, 36.7% also had low levels of education when compared with 12.8% with low levels of education in the urban residence group, $n=144$) [235].

Additionally, a previous review in 2019 investigating the impact of urbanisation on cognitive disorders found that rates of dementia among urban dwellers tended to be lower than those of rural dwellers[312].

In comparison, my research was unable to investigate childhood rural or urban residence, but was able to adjust for intelligence in childhood which could be considered a reflection of education. Additionally, I have adjusted for deprivation (SIMD) and this did not influence the hazard ratio. I have not been able to adjust for deprivation in childhood and this would be important given childhood could be classed as a ‘critical’ or ‘sensitive’ period in life course epidemiology. The SMS1947 is predominantly a survey of Scottish born people of white ethnicity and therefore may not be translatable to other countries, ethnicities and contexts. I discuss this further in [Chapter 7: Discussion](#) in relation to my whole thesis as a key outcome is the need for country specific data, including recognition of diverse ethnicities within different populations, to inform dementia research, policy and clinical care and that data must be representative of the people living in the populations.

Another study has shown, contrary to my results, that rural residence in both childhood and adulthood was associated with a significantly higher risk of dementia at follow-up[236].

A recent (2022) systematic review of dementia prevalence between rural and urban areas in Latin America and Caribbean countries found that dementia prevalence was higher for rural (7.71%) than urban (8.68%) residents [381]. Similarly, a previous review in China in 2014 [693] found that the prevalence of demen-

tia was significantly higher in rural areas than in urban ones (6.05% vs. 4.40%, $p < .001$) [693]. There were similar variations for those with AD (rural = 4.25% vs. urban = 2.44%, $p < .001$), but not for VaD (rural = 1.28% vs. urban = 1.61%, $p = 0.166$) [693]. These authors found that education level may be influencing this as the difference in AD between rural and urban areas was not evident when stratified by educational level [693]. The study also highlighted that the risk factors for AD and VaD differed between rural and urban populations.

In contrast, another recent systematic review (2024) in China found no statistically significant difference between the prevalence of dementia in rural and urban areas [380]. Dementia prevalence in rural areas was 6% (95% CI 4%, 8%; $I^2 = 87\%$, $p < 0.001$), and comparable to the dementia prevalence in urban areas 6% (95% CI 4%, 8%; $I^2 = 98\%$, $p < 0.001$) [380].

My research findings highlighting lower dementia diagnoses in rural areas are consistent with my systematic review findings in Chapter 2 when focusing on High Income Countries (HICs) with moderate to large studies and using routinely collected data. Overall, the systematic review found that rural living was associated with increased odds of dementia, compared to urban living (OR, 1.20, 95% CI 1.03-1.40, $p = 0.0182$; $n = 25,274,741$, $NDementia = 3,462,039$; 33 studies). This was reversed when focusing on HICs with larger studies (more likely to be using routinely collected data as their methods). Whilst there was no effect for HICs between rural and urban living and a dementia diagnosis, there was a difference when considering the size of those studies within HICs (larger HIC studies uses routinely collected data). When further sensitivity analysis was undertaken for the systematic review, my results in this current Chapter are in keeping with what was found on sensitivity analysis in my systematic review in Chapter 2.

In particular, studies conducted in HICs which were moderate-sized ($N > 10,000$) or large ($N > 1,000,000$) showed lower odds of dementia in rural areas (moderate-sized studies: OR 0.93, 0.76-1.13; $n = 25,195,209$, $NDementia = 3,455,479$; 6 stud-

ies; and large studies: OR 0.88, 0.84-0.92; $n=24,900,843$, $NDementia=3,450,748$; 3 studies). Methodologically, the moderate to larger studies in HICs ($n>10,000$) were more likely to be using routinely collected data. If comparing similar methodologies (a key outcome of my systematic review given the heterogeneity in studies: $I^2=99.78\%$) then the findings in my research for this chapter of lower hazard ratios for dementia in rural areas of Scotland (which is a High Income Country and using routinely collected health data methodology) is comparable with what was found in my systematic review.

In summary, my research builds on a growing field investigating location and dementia diagnoses and/or risk. The direction of the association varies: some studies suggesting increased rates of dementia in rural areas and some (like mine) suggesting reduced rates of dementia in rural areas compared with urban. It is a complex area and difficult for direct comparisons with other studies in part due to the key issues I address in this thesis including: definition of place, changes over time in a location, and how translatable the results are between studies and areas if defined differently.

6.4.2 Reasons for differences between rural and urban living and dementia diagnoses

In Chapter 2 Section 2.5.2 I discussed some differences between rural and urban living that might influence life-course risk of dementia. These are pertinent here in considering differences between rural and urban living and dementia diagnoses in my research results, particularly as I do not have multiple time points across the *study* population's lives of their geographical location. Therefore, I will discuss variations broadly that need to be considered that might be introducing bias or error in my findings.

Considering possible reasons for lower rates of dementia in rural areas, a recent study suggested that green space might be a modifiable protective factor against

dementia [286]. This study [286] used a retrospective, population-based cohort assembled from data in the Ontario Population Health and Environment Cohort (ONPHEC) [694] in Canada and investigated urban green space and the risks of dementia and stroke. Whilst this was looking at green space in urban areas rather than rural versus urban areas, the findings could be important for considering possible reasons in my research for lower rates of dementia in rural areas.

Other reasons could include movement at retirement or when care needs increase. For instance, people retiring, particularly if in good health, may be more inclined to uproot from an urban environment to a rural environment for retirement. People who have increased care needs, including people with dementia, may move to a care home near the end of their life. There are more care homes in urban environments than rural environments due to the population size. Finally, people may move to be nearer adult children of working age, and as the urban environment has more jobs this could drive a pull towards urban living in later life. Therefore, there are both ‘push’ and ‘pull’ factors towards living in an urban or rural environment later in life that then may influence (including introducing risk of bias) interpreting dementia diagnoses variation in rural and urban living in this study.

Comparison with SMS1932 study

There are several reasons why my results differ in the direction of association (dementia higher in rural than urban) from research undertaken on the older Scottish Mental Survey 1932, SMS1932, born in 1921 and CHI-linked to routinely collected data in Scotland and variation in dementia (dementia lower in rural versus urban). It is not possible to do a direct comparison due to a number of reasons (different methodologies including different geographical scales, linkage, size of linked data and access to childhood residence). However, it is important to highlight some differences in both of these historical studies at a population level in Scotland. These differences need considered in potentially influencing the different directions of association (higher risk of dementia in rural areas in the

SMS1932 and lower rates of dementia in rural areas in my SMS1947 dementia data linkage study).

In Scotland, the North is less populated than the Southern/central belt. Whilst this does not directly relate to rurality as there are still big cities and towns in Northern Scotland, broadly speaking it is more rural in the North. This is of interest given the past research of the SMS1932 finding dementia rates higher in North (potentially more rural) than South (potentially more urban) Scotland [350]. In contrast, my results showed less dementia rates in rural than urban Scotland. The methodologies of both studies were different with an additional dataset (PIS) incorporated in my linkage which is not solely hospital records (the SMR datasets are hospital records and secondary care data). The geographical scales were also different. Aside from these points, and the difference in results, both studies suggest that the location/place you live is associated with variation in dementia rates, and that this difference may be in adulthood (as Russ et al [350], found no difference in childhood residence and future dementia risk).

Differences in datasets

The research by Russ et. al., linked the SMS1932 to Scottish Morbidity Records (SMR 01 and SMR 04 – general and psychiatric hospital records) and death certificate data (NRS Mortality data) [350]. In contrast my study linked the younger SMS1947 (born 1932) to the same SMR 01 and 04 datasets and the NRS mortality dataset but additionally also to the Prescribing Informations System dataset (PIS). I had additional linkage to PIS for dementia ascertainment. Unlike SMR datasets which are secondary care, the PIS covers community. Whilst neither study could include primary care records from GPs, it is important that the PIS was included in my research to capture those with dementia that might be living in the community and not feature in the secondary care (SMR) datasets.

Differences in geographical locations

Russ et al.2015 used a Bayesian disease mapping model for the geographical

analyses of location and dementia risk and found higher rates of dementia in the north compared with the south of Scotland [350]. They used specific location rather than grouping it into urban or rural residence using the Urban-Rural Classification system, as this was not available at the time of the first Scotland-wide study investigating geographical variation of dementia (SMS1932) [350]. My study methodology differed and was able to use the Urban-Rural Classification system which had been developed in Scotland by the time of my research. I did not use Bayesian disease mapping models. Childhood residence was not available in my linked study.

As I did not want to introduce further bias for those who did not have dementia (and therefore less likely to feature in hospital records from which geographical data was obtained in my study) I took one point in time in adulthood as location for the analysis. In Russ et al.2015, they had traced $n=37,597$ individuals from the original SMS1932 sample of $n=86,520$ [350]. This resulted in an analytical *study* population of 43.5% of the original SMS1932 *source* population compared with my analytical *study* population of 23.9% of the original SMS1947 *source* population ($n=70,805$) [350]. However, my linkage rate to routinely collected data was better as the *database* population had a higher linkage $n=53,037$ (74.9%). The reduction in analytical *study* population in my research was due to missing geographical data. I will discuss how to mitigate this reduction in analytical *study* population in the section on future work below.

In my study, there were more individuals in both groups (those with dementia and those without) in urban regions of Scotland. This is not surprising as the majority of Scotland's population live in urban regions. However, the latest statistics for Scotland suggest that rural living is increasing more than urban living due to more movement from urban to rural areas. Conversely, migration to Scotland over recent years has shown urban areas to have the greatest influx, particularly Glasgow and Edinburgh in the Central belt/South Scotland. This is important for consideration in future research as the demographics of urban and rural areas

in Scotland, and therefore associated risk factors for dementia, may be changing.

Differences in universal healthcare services

In my study the participants in the SMS1947 would have been aged eleven when the National Health Service (NHS) started in Scotland in 1947. The older study participants (SMS1932) of the earlier population-wide linked study investigating geographical variation [350] would have been 26 years old when the NHS started in Scotland. My study participants were mid-life (aged 40s) when the CHI-number system for healthcare linkage started in Scotland in the 1980s. This is contrast to the SMS1932 being in their 60s when it started. This is important as not having access to universal healthcare in the early years and mid-life may have important implications for risk factors for the SMS1932 versus my SMS1947 study participants.

Access to services and differences in healthcare services are important in influencing individuals risk and protective factors for future diseases. Therefore, could differences in my study findings reflect variation geographically in who is diagnosed with dementia? Could those in more remote or rural areas be less likely to receive a diagnosis and this bias affects the results? It could be implied from my results that dementia diagnoses are lower in rural areas. This could be potentially influenced by diagnosis rates varying between rural and urban areas, particularly if a diagnosis of dementia is required by a referral to secondary, or tertiary, care. However, I do not think this is a valid assumption.

In the older SMS1932 study, the authors consider whether different rates of undiagnosed dementia across Scotland influenced their results [350]. They found no variation in rates between Grampian (north of Scotland), 36%, and the Borders (southern Scotland), 44% [350]. In comparison, my results similarly suggest that the proportion of those diagnosed with dementia in rural areas versus urban areas are similar (20.5% in rural and 22% in urban areas).

My results were adjusted for deprivation (SIMD) which was not done in the previous study. This suggests that my results of higher dementia diagnoses in urban areas does not represent deprivation. The *study* population living in urban areas the diagnosis rates are similar in both rural and urban areas in my study.

Access to healthcare services

As highlighted, a difference between the experiences of the individuals in the SMS1932 and SMS1947 surveys is access to universal healthcare, the National Health Service, which started in childhood in the group (SMS1947) I used in my research, but did not in the older SMS1932. This means a direct comparison between the two historical surveys (SMS1932 and SMS1947) *source* populations) is difficult. Across the fifteen years between the two historical surveys, there were substantial changes to access and delivery of healthcare services across Scotland. Therefore, it is difficult to directly show the impact this might have had on an individual's future risk of dementia, but must be considered when comparing the differences in the results and direction of association between rural/urban living and rates of dementia. As discussed in [Chapter 1: Introduction](#) our childhood health impacts on our future health and well-being.

Accessing health services can vary within one country across time, or between different areas. This can be influenced by many factors. It is important to consider this as a reason between differences in rural versus urban settings and dementia diagnoses, in case this reflects access to services - both access to a diagnosis and access to healthcare services for treatment and management. This is important in case of potential bias in the numbers diagnosed with dementia being higher in an urban area with more secondary care services and opportunities for a dementia diagnosis, particularly as anti-dementia drugs are started by a psychiatrist in most areas.

There are known reported variations in dementia care and health utilization for

people living with dementia. A systematic review in 2023 [695] found differences in many dementia care domains (population health, access to services, integration with services, medication safety and patient-centred care) in the 38 included studies between rural and urban areas. People living in rural areas with dementia had: higher mortality rates (Population Health), lower visits to any physicians (Access to services), more hospitalizations but shorter stays (Integration with services), higher antipsychotic medications (Safety of medications), lower use of home care services, and higher use of nursing home (Patient-Centred Care) compared to urban persons with dementia [695]. The authors concluded potential geographical based inequities which need considered in developing more equitable dementia care policies [695].

Last year (2023) a study investigated rural health disparities in health care utilization for dementia in Minnesota [696]. This found some interesting points when investigating three cohorts from 2016-2018 [696]. It found that the rural areas showed 17.6% lower age-adjusted rate (AAR) of dementia mortality than urban areas [696]. The AARs of Emergency Department (ED) visits for dementia were 12.4% higher in rural areas, but this compared with 24.7% lower for AARs of hospitalization for those with dementia [696].

The authors adjusted for income, education, health behaviours within the neighbourhoods, as well as provider access. Following adjustment, the odds ratios of having dementia diagnosis are 12% lower if an ED visit patient lives in rural as opposed to an urban area ($OR = 0.88$, $p < .0001$) [696]. This study highlighted that there are complexities in the differences between those hospitalized or using ED and predictors of having a dementia diagnosis [696]. In summary, the study showed large rural-urban predictors of having a dementia diagnosis through those being hospitalized or visiting ED and the odds ratios of having a dementia diagnosis. It is difficult to understand how this translates to other cultural and country contexts and different healthcare services.

A recent (2024) paper discussed rural-urban differences in the use of health services both before and after dementia diagnosis as part of a population-based retrospective cohort study in Canada from administrative data [366]. In this study, $n=2,024$ individuals ≥ 65 years had dementia identified at index date and a control group of $n=2,024$ without dementia were matched. Rural individuals accounted for around 30% of both cases and controls. In the individuals with dementia, co-morbidity level was lower in rural living compared to urban living individuals ($p<0.05$). There was variation in the number of visits to a family physician in the two years preceding diagnosis and the two to four years post diagnosis, alongside less visits to specialists and lower all-type prescription drug dispensations amongst those living in rural areas than urban area. However, there were no differences between rural-urban living individuals and admission to hospital or short-term institutional care. There may be other factors influencing the differences in Canada's healthcare system with the variations, and lack of understanding of how this would be translatable to other countries.

In summary, there are known variations in literature of diagnosis rates of dementia between rural and urban areas. There are also known variations between rural and urban areas in the access and use of healthcare services (particularly secondary care) that might influence who features in datasets for dementia ascertainment if largely secondary care datasets are used. However, this may be context and country specific (and influenced by other factors) that means it is uncertain how translatable this is to all countries. This is true given there is contrary evidence to suggest some studies have shown no variations.

My study results suggest that the proportion of those diagnosed with dementia in both rural and urban areas (relevant to the *study* population in both those areas) had no difference. Whilst I was not able to investigate differences in accessing healthcare services between rural and urban areas, and limitations of the *study* population being limited to those with geographical data featuring in secondary care datasets at some point in their life, my results do not suggest

that the diagnostic rates varied between rural and urban areas. I was unable to fully explore the variation across the datasets without geographical data also being present in the prescribing information system in my study. In this chapter, Section 6.5 ([Future work](#)), I discuss a way to mitigate any potential bias in the next stages of my research.

Finally, other differences that could be influencing variation in rural and urban areas in my results include the impact of urbanisation on mental health as discussed in [Chapter 1: Introduction](#), which highlighted that the urbanised environment could be more harmful for mental health, including dementia. The world is becoming more urbanised, therefore comparison of cohorts, historical surveys or studies across different decades can be difficult as many aspects of the rural/urban divide have changed.

Deprivation variation between rural and urban areas

In my study I adjusted for deprivation using the SIMD 2004 based on the SIMD decile each individual had recorded on their first SMR datasets entry. There are variations in deprivation between rural and urban areas. It is important to note that not everyone in that area will have the same individual-level of deprivation, but instead that SIMD relates to the whole area.

Of importance, the SIMD can be less helpful in rural areas in capturing deprivation and often reflects a more mixed picture. As highlighted in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), this is because the data zones that comprise the small area assessed in the SIMD are larger in rural areas. As they cover a larger land area, they can reflect a more mixed picture with varying levels of deprivation. Due to this, the SIMD are less able to capture small pockets of deprivation within rural areas compared with the larger pockets in urban areas. The domain indicators for SIMD can still be useful in rural areas if a) combined with other data and b) analysed separately from urban data zones. This was not something I was able to do within my research.

Multiple versions of the SIMD exist. The current version is SIMD 2020 version 2. The version used in the research in this thesis is SIMD 2004. This is because this is closest in chronological age of the Urban-Rural 8-fold classification 2003-2004 used in the research analysis. The SIMD 2004 that I selected was the first edition from the Scottish Government.

By using the SIMD closest year wise to the version of the Urban-Rural 8-fold classification I was using, this was felt to introduce less bias. A recent document (2023) from Public Health Scotland ‘Deprivation guidance for analysts’ discusses the benefits and limitations of using one year’s SIMD for analysis across a longitudinal dataset [668]. The SIMD is compiled through data zones. Data zones change with time. This recent documentation from the Scottish Government discusses some disadvantages of using one SIMD for longitudinal analysis [668]. This includes: *"The further you get from year of the population and indicators used in the particular SIMD release you have chosen, the less ‘pure’ your deprivation deciles/quintiles are likely to be."* [668]

This document also highlights that data zones can improve or worsen over time, e.g. a data zone that is in the most deprived category for one release could have moved to the least deprived category for another (and vice versa) [668]. This could be referred to drift over time. The report highlights an example of this that 33% of the Data Zones in Glasgow City Council area were in the 5% most deprived in Scotland in SIMD 2004, which then fell to 24% in SIMD 2006 and 23% in SIMD 2009 [668]. Therefore, if a single SIMD release, such as I used SIMD 2004, is used to identify the most (and least) deprived Data Zones, then it will not be able to reflect changes over time. Whilst this is a limitation, for my study it also reduced introducing further bias in not having multiple time points for everyone to compare geographical location and deprivation (SIMD) at different time points.

6.4.3 Strengths, limitations and challenges

In this section I discuss the strengths limitations and some key challenges of this research study.

Strengths:

Population size and data linkage

A major strength of this study is the ability to link individual level data from a historical population-wide national survey, the Scottish Mental Survey 1947, (which includes a childhood intelligence variable) to routinely collected data (health and mortality datasets) at a population level over 60 years later. The Scottish Mental Survey is unique in having documented childhood intelligence, and long term follow up to an age (80 years) when dementia prevalence is high enough to consider life-course predictors. Data linkage was possible due to Scotland's NSH which can comply with the current GDPR, data governance and the Five-Safes Framework as a recognised Trusted Research Environment (TRE). Additionally, the unique healthcare identifier in Scotland, the CHI-number, allowed linkage between several data sources to maximise the ascertainment of a dementia diagnosis.

Another significant strength is the ability to obtain very good data linkage of an external population-wide historical study (SMS1947) with individual data to routinely collected health data. The data linkage from *source* population (SMS1947) to *database* population (SMS1947 CHI-linked to routinely collected data) was approximately 75%. It is inevitable that linking additional datasets, particularly historic ones, result in loss of some individuals in the linkage process (e.g. some may have left Scotland or died prior to the introduction of the CHI system or changed their surname and be harder to trace).

Despite this, my study had significantly more data linkage from *source* to *database* population than the older SMS1932 study, which had a linkage for this stage of under 50%. This meant a strength of my study was being able to create my

study population from a large *database* population. My *study* population size of $n=16,910$ was large. The strengths of the results obtained are that they are reflective of a large sample size, and also are representative of a historical population sample from 1947.

Childhood intelligence

A further substantial strength is having childhood intelligence as a variable. It is known that childhood intelligence can impact on future dementia risk [677, 678]. It is rare to have data on childhood intelligence that can be linked as a study to routinely collected data over 70 years later, and therefore a major strength of this study. By having childhood intelligence as a variable it allows for adjustments in statistical analysis. Childhood intelligence would otherwise be a potential confounder that could not be adjusted for.

Although childhood intelligence is not directly comparable to low education, it could be sometimes considered a marker for future education attainment. Therefore, importantly my results of rural versus urban living were not influenced by adjusting for childhood intelligence. If extrapolating this then it could be proposed that they also do not reflect low education in childhood as a reason for the variation in rural and urban residence in adulthood and dementia diagnosis in my study.

Clinical experience

A strength of this study is my clinical experience of understanding the recording of dementia in hospital datasets and death certificates. It has helped me understand potential bias and limitations of the datasets and recording of dementia that might impact on the interpretation of the results.

Geographical scale for rural and urban areas

This study includes a geographical scale (the Urban-Rural Classification System) that is specific to Scotland [663], validated, and created to be used over time.

This is a strength, particularly given the complexity of defining rural and urban areas, that has been highlighted in my thesis in [Chapter 1: Introduction](#) and [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#). This allows consistency in the rural/urban variable in the analysis, and also allows further analysis of whether those rural areas are remote and accessible or not.

One challenge of any geographical scale is the consistency over time. This was highlighted in a previous paper discussing the impact that changing configuration of census zones for each UK census has on small geographical area analysis [697]. Those researchers suggested three sets of consistent areas through time (CATTs) that could be used from 1981-2001 Censuses [697]. Whilst these are sometimes still used in longitudinal analysis based on the Carstairs index, the Scottish Government has recommended they are not used for analysis from 2011 onwards [668]. Therefore, I used the Urban-Rural Classification System which was designed specifically to consider geographical scales over time, and reduce introducing bias with one scale across all analysis.

Of the many versions of the Urban-Rural Classification system, the Urban-Rural Classification for 2003-2004 was proposed to be the most consistent over time which is why in the Methods I use this specific version for the analysis of rural and urban areas. The Urban-Rural 8-fold 2003-2004 classification can be used across different years, but is based on the 2003-2004 classification.

Limitations:

I will discuss some important limitations of my research. These include the size of the *study* population, missing geographical data, selection of one SIMD and Urban-Rural classification for a longitudinal time period, and extracting dementia diagnoses from mostly secondary care data. Some of these limitations are pertinent to routinely collected data health research in general and some have important considerations for future research specific to my research question.

Missing geographical data

A major limitation of this study, were missing geographical data (rural/urban living) data which reduced the size of the *study* population from the *source* population. At the time of my research, only the SMR 01 and SMR 04 datasets had geographical data available in the NSH linked with the project. Therefore, this meant that geographical data was not available for everyone in the *database* population (missing geographical data in the *database* population was $n=36,127$, 68.1%).

As a result, the *study* population (those within the *database* population with geographical data available) was a substantially smaller sized population ($n=16,910$), representing 31.9% of the total *database* (CHI-linked SMS1947) population. This resulted in fewer dementia cases that could be included in the *study* population as approximately 1,000 persons with dementia in the *database* population could not be included in the final study population due to missing geographical data (Urban-Rural classification variable).

The geographical data was essential for analysis. So whilst this study's limitation includes a reduction from *database* to *study* population size, it did not reduce the ascertainment of dementia as significantly as a proportion of the *study* population. There is the potential in the *database* population and *study* population having the risk of bias as the geographical data was limited to those who had hospital admissions (SMR datasets). I discuss this further in Section 6.5 ([Future work](#)) how this could be further addressed.

Rural population size and structure

The population size of this most rural/remote regions of Scotland was $n=770$. Whilst the HR are below 1 in my results, indicating that dementia is less likely in rural areas, some of the CIs cross 1 (e.g. the 'rural-remote' area had $HR < 1$ with a narrow consistent range from 0.81-0.84) likely due to small numbers. However,

when all rural areas are pooled together to a larger population size ($n=2,563$) then the 95% Confidence intervals were slightly narrower (0.77-0.92).

A limitation, and potential risk of bias, is the validity of SIMD in rural areas as a marker of socio-economic status. SIMD is useful in reflecting the level of socio-economic status for an area, but not individuals within that area. However, as mentioned in Chapter 5, there are limitations with SIMD in rural areas. As the data zones used for SIMD are much larger in rural areas, this can make it harder to ascertain deprivation in rural areas. Whilst, I have adjusted for SIMD between rural and urban areas, a limitation is that I cannot exclude a difference (and possible risk of bias) in deprivation between rural and urban areas.

Predominantly secondary care datasets for dementia ascertainment

This might create bias as those without dementia *may* be the ‘healthier’ group and therefore restricting this ‘control’ group to a secondary healthcare dataset may influence the results. From the ***database*** population ($n=53,037$), there were $n=36,127$ (68.9%) individuals without geographical data who had to be removed to create the study population ($n=16,910$). Geographical data was essential to the research question. Imputing geographical data, as can be done for other missing data, would have introduced potential significant bias as the variable was binary, not continuous, and the key predictor variable.

Ascertainment of dementia in the ***database*** population was $n=4,767$ (9.0%), and in the ***study*** population was $n=3,676$ (21.7%). This was a decrease in 1091 people (22.9% reduction) with dementia between the two populations. There was a significant drop in ascertainment rate of dementia from the whole ***database*** population to the ***study*** population.

There could be many reasons for this, but one reason is due to the geographical data only being available in the SMR (hospital/secondary care) datasets in my project in the National Safe Haven at the time of my research. It is important

to consider an individual's journey with dementia through different healthcare settings to understand where dementia will, or possibly not, be recorded.

For instance, there could be many more either dying with dementia recorded on their death certificate, or receiving anti-dementia drugs on the prescribing dataset that are also not admitted to hospital at some point in their life-time. They may not feature in my *study* population. There is also the potential that there were individuals who were prescribed anti-dementia drugs in the community and therefore found via the PIS database in the *database* population, but if they then stopped that anti-dementia drug (due to side effects, patient wishes or not appropriate at the end of life) and were subsequently admitted to hospital there is potential that if their admission reason was not due to dementia that they could be coded for other discharge diagnoses and dementia not included.

If someone was admitted to hospital and had dementia (but not on anti-dementia drugs) and subsequently died in hospital there is a reliance on ensuring that secondary care doctors include dementia on the death certificate. If they felt it was not contributing to the acute cause of death this may not be entered. The individual would feature in my *study* population as they had entered SMR datasets (hospital admission) but would not definitely have an ICD code on their death discharge notice nor death certificate.

The lack of primary care data, specifically GP data, is a limitation in dementia ascertainment but also ascertaining information on the 'control' group as the potentially 'more well' group they may not be interacting at all with secondary healthcare data (SMR 01 and 04) and without GP data may not have geographical data available for the analysis creating bias. The CHI-number has the address of a patient as recorded in their primary care records and for my study this was available only in the SMR datasets (secondary healthcare data). Not everyone with dementia is admitted to hospital, receives dementia drugs or has dementia recorded on their death certificate. It is known that if dementia is recorded in

routinely collected data then there is a high positive predictive value that that diagnosis is accurate [152]. However, without primary care data from GPs there will be individuals that could have been missed with a diagnosis of dementia that don't feature in secondary care datasets (SMR 01, SMR 04) or PIS or NRS Mortality datasets.

My dementia ascertainment potentially reflects those most interacting with these health and administrative datasets and not all dementia. People with dementia are more likely than those without dementia to be admitted to hospital [698] (and therefore have an SMR record) or interact with healthcare services (e.g., prescribed a dementia drug in PIS dataset). However, this might be only a theoretical limitation. By age 80 there is a probability of 1 in 6 chance of dementia in the population [567, 670]. Therefore, in my study population of $n=16,910$ with an end censorship of 2018 (when those alive would have been in their 80s) it would be anticipated to expect around $n= 2,818$ people with dementia in the study population. My results are comparable with what would be expected for the study population size. Therefore, despite not having GP records, the dementia ascertainment is above what would be expected.

Interaction with healthcare services will vary between those who are healthy and those who have illness. The datasets in this research were not all secondary healthcare as the PIS dataset is used in the community/primary care as is the NRS deaths dataset. Whilst there is no ability to make an assumption that this definitely resulted in bias, it is important to consider for future work to minimise any bias. Much of this was outside my control at the time of the analysis due to the access, and permission, to use geographical data. A recognised limitation with all routinely collected healthcare data that uses largely secondary care data and not primary care data (such as GP data) means that those without dementia may not feature in secondary care healthcare datasets as frequently.

I discuss separately in the Section 6.5 how to try and mitigate this limitation

in future studies. It highlights what is already known within routinely collected health data, which is that as many datasets are needed as possible to accurately ascertain dementia [518]. This is challenging with separation still between primary and secondary care routinely collected health datasets.

In summary, this significant drop of dementia ascertainment from *database* population to *study* population could represent many people living well with dementia in the community who are not admitted to hospital, or choose to be cared for in the community if they have severe dementia and it reflects holistically their past wishes, their family and health/care teams. Secondly, there could be those with dementia admitted but for reasons do not have it coded on their discharge. Given that 21.7% of the *study* population were diagnosed with dementia, and 8.9% of the *database* population were diagnosed with dementia, this higher prevalence suggests that the *study* population group were potentially a sub-group of the *database* population who were more frail. There could be ascertainment bias within the study population in interpreting the rural/urban living given that geographical data was only available in hospital (SMR) records. In Section 6.5 (Future work) I discuss this further.

Geographical and deprivation scales across time

Geographical and deprivation scales change across time. The Urban-Rural Classification system is for a specific year but can be applied across different time periods [663]. The SIMD (used for deprivation) however is specifically for a year. As highlighted, in 6 Reasons for differences between rural and urban living and dementia diagnoses I used SIMD 2004 as this was closest in chronological age to the Urban-Rural 8-fold classification 2003-2004 used in the research analysis, to reduce introducing bias. However, the SIMD uses data zones to comprise the SIMD. Data zones can change over time (one data zone becoming less or more deprived) and so using a SIMD for one time period will not capture this. It also does not highlight deprivation areas in rural areas as accurately as urban areas. These are limitations, but using multiple SIMD versions when there were not con-

sistent geographical points between the two groups (those with and those without dementia) could have introduced further bias in analysing and interpreting the results.

The nature of the routinely collected data and privacies (and ensuring individuals were not identifiable) means there was not the ability to mark an individual's geographical location and also SIMD year by year during the years until 2018. For those it was possible to do this and gather multiple geographical locations across their lifetime according to the 8-fold classification, this would be limited to only those who had interacted with the SMR datasets so introducing further bias. Additionally, for those without dementia, it would be difficult to justify at what point you were to collect geographical and SIMD data for these 'controls' versus those with dementia ascertained. Whilst using one SIMD year for all adjustments in the CPH models is a limitation, it would introduce further bias using different SIMD (or geographical scoring) across the time-frame. This bias would be increased as it was not possible to be clearly able to define the 'controls' across the routinely collected data time period. Due to the nature of the 'controls' not having dementia, it could mean they may interact less with healthcare datasets introducing increased bias in the analysis.

In summary, areas change over time, and individuals can change the areas they live in over their life-time. It was not possible to obtain multiple time points to consider the timing of the exposure (rural/urban living) and dementia outcome, nor consider a 'cumulative' or 'sensitive' risk across an individual's life time, in a life-course epidemiological perspective.

Challenges

There are common and frequent challenges in using routinely collected data for secondary use in healthcare research, and there has been rapid development in this field during the course of my PhD. Therefore, I will also discuss some of these challenges in this context.

Timely access to linked data

The first challenge was the delay in requiring the data in the original plans for the study, this is outlined in Chapter 5 Table 5.6 ([Timeline of planning SMS1947 research in Scotland](#)). This was due to the data requiring re-linked due to the loss of a data analyst within the data linkage team whom was aware of where all the individual linked files were. This resulted in around an eighteen month lag time from approval for my project within the first study and being allowed to access the data within Scotland's National Safe Haven. During this time, the study project came to an end and although an extension was sought, and granted, it became clear that it was not appropriate to continue on an older CHI-linked data set of the SMS1947 cohort as a new linked study was being created (with permissions and ethics sought) with more recent data linkage.

Timely access to linked routinely collected health data for healthcare research is a common issue but less evident until working with this type of data and awareness in variations in the permissions and approvals processes. Routinely collected healthcare data is generated frequently, often daily. However, it is still rare, and difficult, to be able to access that data as a researcher in a real-world 'live' interaction. This is because there are rightly processes, permissions and stages in place to access the data and additionally linkage of data by data scientists can take time. In Chapter 5 I discussed the timescale for undertaking this research, including different interactions in the planning to access linked SMS1947 study data to routinely collected data.

Processes and permissions are essential to ensure the Five-Safes Framework, GDPR compliance, and appropriate secondary use of routinely collected data. However, they can also culminate in delays in accessing data for the researcher to analyse the data. It is a challenge in balancing these stages whilst also ensuring timely access to data so that results can inform new research directions or influence changes for those living with a disease. There are differences in how data is

linked which can either increase the access speed for the researcher or delay the time.

In many ways I was fortunate that my PhD required separate extensions, outside of this work, due to maternity leave, that afforded me the opportunity to work on the new study. Had that not been the case, I would have lost a lot of time unable to undertake novel research due to the data linkage issues. There was a lot of development and learning through this experience and I feel this challenge has really shaped my understanding of different processes in data linkage at a national scale – particularly when I later undertook the research in ANZ (Chapter 4) which has a different approach to accessing and linking data.

Age of diagnosis of dementia

One important challenge is the recognition that the coding or mention of dementia in routinely collected data does not necessarily represent the age of diagnosis. It is not possible to read the corresponding outpatient or inpatient summary from SMR to acknowledge whether this was a new diagnosis or not. Therefore, it is a common assumption, and acceptance, in routinely collected data healthcare research that the first mention of a diagnosis is analysed as the first perceived data of diagnosis. If unable to include primary care data this assumption may not be correct if a diagnosis is first recorded in primary care datasets rather than secondary care. Using routinely collected data for healthcare research is reliant on a) accurate coding of medical records and b) interaction with secondary healthcare datasets if no primary care/GP records are part of the research.

There are movements to further integrate primary and secondary care datasets to mitigate this challenge. Moving from different coding systems in primary care (READ) and secondary care (ICD codes) to an integrated system (SNOMED CT) [548, 475] will be valuable for gathering holistic information across different care settings by using structured clinical vocabulary that can cross different datasets. Additionally, the role of Artificial Intelligence (AI) is increasingly being deployed

as there is a lot of valuable free text data in the ‘notes’ in primary care records. Until linked data between primary and secondary care, alongside additional resources such as social care and the ability with AI to interrogate written data improves, then this is a recognised challenge and limitation.

Researcher support

It is the nature of the system that research co-coordinators within eDRIS change over time, this is more of a challenge if there isn’t timely access to linked data. There is a lot of information the research co-coordinators have through the multiple discussions they have with researchers. Whilst this is captured in written format, it does not negate the benefit of having one consistent research co-coordinator throughout the whole project who is familiar with the project in detail. I had a number of research co-coordinators in eDRIS during my PhD, at one point several in short succession, which made it challenging for them to become familiar with a project in a short space of time and for me too. Therefore, it is positive to see this year that Research Data Scotland (RDS) create the Researcher Access Service which is a streamlined process and approvals process which promises to improved timely access to some datasets, continuity in processes and support [553].

Shared learning amongst health data researchers

Lastly, I feel that many of the researchers have similar difficulties at times but there is no shared secure platform to discuss these within the National Safe Haven in Scotland. Within the ANZ IDI there is a safe and secure platform to discuss common coding or linkage issues. Individuals can post issues with statistical coding or difficulties with extracting different ICD codes and share learning within a secure environment meeting the Five-Safes Framework. However, the nature of the set up of Scotland’s NSH does not currently allow this. There can be missed opportunities for shared learning, and a tendency to work in silos or be dependent on co-researchers for your specific project who may or may not have encountered the same issues before. There is always the escalation of questions and IT issues

via the eDRIS research coordinators but I feel there is still a need for a separate way of shared learning of researchers doing different projects within Scotland's NSH.

These are some of the challenges as a researcher using a National Safe Haven. Many of these challenges are common across different Safe Havens (Trusted Research Environments, TREs) [498]. For this study, my research took place during some significant updates in Scotland's NSH from Version 1.0 to Version 2.0. This resulted in additional 'outage' days when the IT made the transition to Version 1.0 to Version 2.0. From a researcher's perspective, there were changes in accessing the system (which is via a remote Virtual Project Management, VPM, system) with password and login changes and how projects were saved within the NSH.

Over the course of my PhD I became more aware of these minor challenges, and being able to anticipate how to navigate them in planning work. Technical issues can develop at any point and when using a NSH, a researcher is reliant on the supporting IT systems for the NSH as it is part of a controlled network. One example of this is updating R studio statistical programming within the Scottish NSH. This is typically a web-based system, and programmes (packages) are frequently amended in an open-source platform and available via R's CRAN server [699].

As there is no access to the internet within the NSH (for security reasons as it contains sensitive research data), it means that the researcher is unable to update the R studio themselves in the timely fashion as they would on a web-based access set up. Most of the error codes are simply warnings of the R studio version being older than the packages being used in the R studio. The package and function still run, but it does mean extra checks to ensure whether a newer package is required. It is the nature of R Studio that you must always check for compatibility issues or updates (easily done via a web-based R studio) in the R

packages being used. This is harder when working in a non-web-based system in the NSH reliant on updates across the system rather than a researcher being able to update the software.

6.5 Future work

Linking a historical study (such as the SMS1947) with routinely collected health data over seven decades later for healthcare research is a unique opportunity. It allows inclusion of historical population-level data for 1947 (a period of historically great change including development of the NHS) with childhood intelligence to be linked with recent electronic health records. It allows a longitudinal perspective and a ‘virtual’ follow-up via routinely collected data to ascertain dementia anonymously (in line with data governance and the Five-Safes Framework). It also allows the inclusion of all people, irrespective of their ability to consent to research as it is anonymous and de-identified.

This is important as often those with dementia have been excluded from some research, but it is vital that research is undertaken to support and improve the lives of those living with dementia. Having childhood intelligence allows adjustment for a known variable that can influence future dementia risk. This approach is a valuable research methodology, albeit with some limitations as outline above. Therefore, I will now discuss how future work could 1) mitigate some of these limitations, and 2) further investigate the causes and explanations for the difference in dementia prevalence between urban and rural areas.

Expanding geographical data

Firstly, a key limitation was access to geographical data across all the datasets (SMR, PIS, NRS mortality) within the database population in the National Safe Haven. For my study, only SMR had geographical data, and future studies should include geographical information from PIS and NRS mortality. I have submitted, and had approved, a request to do this to HSC-PBPP as a minor amendment

for permission and access to geographical data in the PIS and NRS mortality datasets, to create an expanded study population. This will increase the study population size (with less lost from the database to study population stage). Approvals were not granted in time for the release of the data to be included in this thesis.

Once the study population has been created with geographical data from all datasets then I will aim (if analysis allows) to extract three time points for each individual using the Urban-Rural 8-fold Classification 2003-2004 system. These time points will include: the earliest date and corresponding Urban-Rural 8-fold Classification 2003-2004; the last date and corresponding Urban-Rural 8-fold Classification 2003-2004; and finally a mid point between the two with the corresponding Urban-Rural 8-fold Classification 2003-2004. This will be obtained for both those with and those without dementia. It will then allow me to investigate rural or urban living across the time period from 1980s (when the CHI system was first developed and routinely collected data started) to 2018 (end of censorship).

This should allow the ability to consider whether there is a ‘cumulative risk’ of living in a rural or urban area and dementia or if there is a ‘sensitive’ period during a certain age period that living in a rural or urban environment has greatest influence on future dementia risk. Childhood residence will not be available in this new database population. I mentioned in [Chapter 3: Using big data for dementia research](#), there is the potential that a unique identifier for geographical data such as the UPRN [379] may in the future provide the opportunity to look at this whilst minimising the risks of identifying any individuals. It may also then be possible to link UPRN with other environmental factors such as air pollution data nationally, sunlight exposure.

Considering variables driving rural/urban differences

Secondly, future work needs to focus on understanding what might be driving the differences between urban and rural areas to explain the difference in demen-

tia outcome. Whilst we do not know definitively the diagnosis rate across rural and urban areas of Scotland compared with the anticipated dementia rate, other studies in HICs would suggest that rurality does not influence dementia ascertainment. This would be supported by my results suggesting that the prevalence in rural versus urban areas of dementia were similar (22% of residents had dementia in urban areas versus 20.5% of residents with dementia in rural areas).

The focus for future work needs to examine factors across the life-course, but particularly relating to adulthood location and future dementia risk. This is because the previous SMS 1932 study showed no difference between childhood residence and future dementia risk. It did show a difference in adulthood residence and dementia risk, which my results also showed (although with a different direction). The focus could be in the time period from age twelve to forty, or from forty years onwards. There are differences between the two studies of SMS1932 (rural = higher risk) and SMS1947 (urban = higher risk): the NHS started in 1947 so from age age onwards the second SMS group had free healthcare, and also access to free dental care. Another difference is that children in the SMS1947 study all experienced war rationing from birth to age eleven years. Despite the severity of war rationing, children had more generous ration books than adults and also had regular supplementations (e.g. cod liver oil). This population had potentially better or more consistent nutrition across all levels of deprivation and urban/rural living at ages when they were growing compared to the older SMS1932 group. This could influence differences in my results between the two groups, as the first SMS1932 did not have a free national health service until they were twenty-six years old. It is known that childhood health and well-being impacts on later life morbidity and mortality.

One approach would be starting to focus on those aged 40 years onwards and risk factors, as this is the age when the routinely collected data started in my younger SMS 1947 study. This could be done using an additional cohort to investigate possible risk or modifiable risk factors that might influence dementia

as a representative of an ‘urban’ or ‘rural’ group. The Lothian Birth Cohort 1936 (LBC1936) would be an ideal study to further investigate this in. People who had taken part in the SMS1947, and were living in Lothian, were approached in 2004 to take part in a detailed study of their cognitive and physical health, and the seventh wave of data collection has started in 2024. Their data have been CHI-linked to routinely collected health data but also have additional face to face assessments and data collected including many risk factors for dementia, and detailed history of their childhood residence.

The Lothian region is largely considered an urban area in the Urban-Rural 8-fold classification system. Given my results suggest a higher Hazard Ratio of dementia in urban areas, this would be an opportunity to explore some of the possible risk (or modifiable/protective) factors within a sub-set cohort of the SMS1947 exploring factors over time that might be influencing this increased risk in urban settings, with detailed life-course data. This could include modelling changes in risk factors such as air pollution over time in the Lothian region as well as including adjustment for the outcome (dementia) according to dementia risk-factor indices and cumulative rural/urban living environment or other factors due to environmental living such as deprivation over time, as well as genetic risk factors, lifestyle and social factors (e.g. exercise, social connectedness). The data collected on the LBC1936 cohort would allow a more detailed deeper investigation of understanding this finding of urban living and dementia risk. Within the LBC1936 cohort it would be possible to adjust for childhood residence of rural or urban living on future dementia risk.

To allow comparison with a more rural cohort, and comparison between the North and South of Scotland, the Aberdeen Birth Cohort 1936 (ABC1936) could be used. Individuals in the LBC1936 and ABC1936 may not all live in the cities in those regions and so there may be a level of variation within the Urban-Rural 8-fold Classification 2003-2004 to allow investigation of differences between broadly rural or urban areas within the classification.

Another approach would be to use the LBC1936 cohort and consider population level risk factors at different points across the life-course influencing dementia as outlined in the 2024 Lancet Commission on Dementia Prevention [97]. As it is observational it would not demonstrate causation but it might help explain possible variables driving the increased rates of dementia in urban areas. These results might generate hypotheses to further investigate through a causal pathway.

6.6 Chapter conclusions

In summary, my research investigated the question: *“Is there a difference in dementia rates between rural and urban areas of Scotland in the historical SMS1947 CHI-linked to routinely collected health data.”* I was able to use multiple routinely collected health datasets CHI-linked with a historical population-wide survey with individual-analysis to investigate dementia in those over 80 years old in the survey and compare rural and urban living with a dementia outcome.

My results found that rural living (Classes 6-8) was associated with a 16% (CI 8-23%) lower rate (lower hazard ratio) of dementia than urban living. Dementia diagnoses were lower in rural areas versus urban areas, with lower Hazard Ratio of a dementia diagnosis if living in an rural area (unadjusted HR 0.84, 95%CI 0.77-0.92 and adjusted for sex, intelligence and SIMD, HR 0.84, 95% CI 0.77-0.93) compared with urban areas. Further analysis showed that the HR was lower if living within the most rural/remote region of Scotland (unadjusted HR 0.77, 95% CI 0.61-0.96, and adjusted for sex, intelligence and SIMD, HR 0.77, 95%CI 0.61-0.98). These results were not altered by sex, intelligence, SIMD when adjusted in different statistical models. This is important given there are known links sex (female) and increased risk of dementia, known links between deprivation (SIMD) and health inequalities including dementia risk, and known links between lower intelligence in childhood future dementia risk.

The strengths of this research included using a large population-wide historical dataset with individual-level data linked to routinely collected health data, the size of the population, multiple datasets to ascertain dementia, and a universal rural/urban classification system unique to Scotland. Additional strengths were the ability to adjust for childhood IQ and later life deprivation.

The limitations of this research include missing geographical data from two of the three routinely collected health datasets, limiting the size of the *study* population from the original CHI-linked *database* population. This also limited the ability to investigate geographical location (rural/urban) living across a person's life-course.

Future work has begun investigating this research question with the increased geographical data in other (PIS and NRS mortality) datasets, rather than solely the SMR datasets to reduce potential bias in the *study* population. Future advancements in a unique identifier for geographical data (UPRN) [379] in the UK could prove promising for expanding research in this area. I have highlighted how further work could investigate the driving factors explaining the differences between rural and urban areas.

Chapter 7: Discussion

7.1 Chapter introduction

In this discussion I draw together my findings and key points from each chapter. I also discuss my thesis overall with its implications and suggestions for future research in the context of my research question, *‘Does dementia vary between rural and urban areas, and how can we use routinely collected health data to answer this question?’*.

There is a growing interest in the role of ‘location’ in influencing an individual’s health, and risk factors for future development of dementia. Therefore, in this thesis, first reviewed the recent literature systematically, and analysed rural-urban variation in relation to dementia ([Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#)). I then considered the re-purposing of routinely collected health data for healthcare research with a focus on its use for dementia and to investigate geographical variation of dementia ([Chapter 3: Using big data for dementia research](#)).

Following this, I undertook novel research in Aotearoa New Zealand (ANZ) ascertaining dementia from routinely collected data and considering the geographical variation between North and South Island ([Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#)). Finally, I investigated rural/urban variation and dementia in the historical Scottish Mental Survey 1947 (SMS1947) linked to routinely collected health data ([Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), and [Chapter 6: Results of the SMS1947 dementia data linkage study](#)) to examine if rural or urban living was associated with differences in dementia diagnoses rates.

The motivation for undertaking this research was to contribute to a developing important field of dementia epidemiology, to consider how to investigate this area

using routinely collected health data, and look at additional potentially modifiable risk factors to reduce the risk of dementia to improve our understanding of dementia.

7.2 Summary of findings

In [Chapter 1: Introduction](#), I outlined my overall thesis topics, my aims and objectives. I gave an overview of dementia discussing how it can be defined, a brief overview of pathophysiology and genetics and how it can be diagnosed. I then discussed epidemiology in relation to dementia focusing on dementia life-course epidemiology. I finished by discussing an overview of geographical variation of dementia with a focus on the terms health and medical geography, geographical determinants of brain health and models that consider geography in relation to ageing and dementia risk.

In [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#), I presented a systematic review with meta-analysis of observational studies investigating dementia risk with rural versus urban living. Of $n=13,053$ records identified, $n=545$ full-text articles were assessed for eligibility, $n=38$ papers were included in the systematic review, of which $n=27$ papers could be included in a meta-analysis. Overall, rural living was associated with increased odds of dementia, compared to urban living (OR= 1.20, 95%CI 1.03-1.40, $p=0.018$).

However, when stratifying by country income, this pattern was seen in Low-Middle Income Countries (L-MIC), (OR = 1.33, 95% CI 1.06-1.68, $p= 0.014$), but in High Income Countries (HICs) there were lower odds of dementia in rural areas compared to urban (OR=1.01, 95% CI 0.88-1.16), $p=0.884$). When further stratifying by large studies (using routinely collected linked data) in HICs, the direction was reversed with higher risk of dementia in urban rather than rural areas (OR 0.88, 95% 0.84-0.92).

These results build on past research and expands discussion on how we classify rural and urban within and between countries. Future research should consider heterogeneity in the definitions of ‘rural’ versus ‘urban’, as well as growing disparity in methodologies between HIC (large routinely collected data studies) and L-MIC (traditional smaller face-to-face epidemiological studies).

In [Chapter 3: Using big data for dementia research](#), I discussed how big data, in the form of routinely collected health data, could be repurposed for health-care research, specifically for dementia research. I presented my peer-reviewed publication on big data in Scotland. I highlighted the process for using routinely collected health data for research including key concepts such as The Five-Safes Framework, Data Governance, unique identifiers and data linkage. Finally, I consider how future ageing health research using routinely collected health data could expand in the UK.

In the following Chapters ([4](#), [5](#) and [6](#)) I used knowledge and learning from the initial chapters to subsequently undertake novel research in Aotearoa New Zealand (ANZ), and also Scotland. I used routinely collected health data to investigate variation in dementia geographically at different scales.

In [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#), I investigated the variation of dementia rates between North and South Island in ANZ using a large national data resource, the Integrated Data Infrastructure (IDI), comprising of many routinely collected datasets. The prevalence of dementia in the ≥ 60 year olds varied including between North and South Island (e.g., for 2012, North= 9,696/637,245 people [1.52%], and South= 3,702/230,001 people [1.61%]). The overall total percentage of people with dementia remained relatively consistent across the years studied (2012-2015) in both Islands. People with dementia were approximately 0.3% of ANZ’s total population in 2015. Most cases were ‘unspecified dementia’ sub-type, and around 9%

had more than one dementia sub-type. Most dementia was amongst the 80-89 age group.

Dementia counts were highest amongst the European ethnicity, but when age and sex standardised, dementia cases were higher for Māori than non-Māori (2% vs.1%). These results were cautious estimates, with the potential for ascertainment bias, and are lower than anticipated which again highlighted the urgent need for a larger national epidemiological study in ANZ using ANZ specific data. My methods have since been used by local and international research collaborations investigating dementia in ANZ, and expanded to include additional datasets to increase dementia ascertainment.

In [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), I outlined the process to undertake a research project investigating dementia variation in a historical national Scottish survey (the Scottish Mental Survey 1947, SMS1947) CHI-linked to routinely collected health data. The focus was on investigating rural and urban variation and dementia rates (dementia diagnoses) within this *SMS1947 dementia data linkage study*. This Chapter highlighted the stages to develop this project, challenges and discussed the methodology used.

In [Chapter 6: Results of the SMS1947 dementia data linkage study](#), I reported the results from my novel research investigating geographical variation (focusing on rural and urban living), and dementia diagnoses in the Scottish Mental Survey 1947 survey CHI-linked to routinely collected health data (*the SMS1947 dementia data linkage study*). Overall, the results showed that living in a rural region was linked with lower hazard ratio (unadjusted HR 0.84, 95%CI 0.77-0.92 and adjusted for sex, intelligence and SIMD for deprivation, HR 0.84, 95% CI 0.77-0.93) of developing dementia than living in an urban area. Further analysis showed that the HR was lower if living within the most rural/remote region of Scotland (unadjusted HR 0.77, 95% CI 0.61-0.96, and adjusted for sex, intelligence and SIMD, HR 0.77, 95% CI 0.61-0.98).

In summary, the thesis findings clearly highlight that variation exists between the location a person lives (rural or urban) and their subsequent development of dementia. The underlying factors driving this difference, and how this varies between HIC and L-MICs, are not fully understood. This thesis has shown an association between geographical location and dementia diagnosis, but this does not imply that this association is causal. This highlights important aspects for future work including how a place is defined as rural and urban, the diagnosis of dementia, how this translated to different contexts, countries and ethnicities, and a need for future work to understand the potential causes for this association.

My thesis highlights the developing work considering how to integrate geographical data within routinely collected datasets so that individuals are not identifiable, and that when I started the PhD there was no universal agreed definition of ‘rural’ and ‘urban.’ Whilst this is changing, there are still difficulties for incorporating geographical data into routinely collected health data for meaningful analysis.

7.3 Strengths, limitations and challenges

In Chapters 2, 4 and 6 of the thesis, I have highlighted the strengths, limitations and challenges of the studies. Additionally, in [Chapter 1: Introduction](#) I discussed the challenges of diagnosing dementia clinically and in [Chapter 3: Using big data for dementia research](#) how this impacts on ascertaining dementia from routinely collected health data along with general discussion of the strengths, limitations and challenges of re-purposing routinely collected health data for research. In this section I aim to extract and highlight some of the common themes in relation to the strengths, limitations and challenges throughout the thesis.

Strengths:

Overall, a main strength of the thesis is the ability to have large population sizes in the analysis through the use of large routinely collected datasets at a population-

level but with individual-level information in both Scotland and ANZ. Therefore, the results from these studies are not from small sample sizes with increased potential for bias in their interpretation, although it is important to recognise that bias can also increase with larger sample sizes.

A major strength of the *SMS1947 dementia data linkage study* in Chapter 6 is the ability to link a national historical survey (which has a variable for a childhood intelligence score to allow adjustment for childhood intelligence as a confounder) in Scotland to routinely collected health data. This provided a population level survey from 1947 with individual-level analysis *nested* within routinely collected health data to investigate my research question. My results from this research were in keeping with the subset analysis of the meta-analysis from the systematic review in Chapter 2 which showed increased dementia in urban versus rural areas in HICs using routinely collected data.

Another strength of this thesis is the ability to use multiple datasets for ascertaining dementia. This included prescription dataset in Scotland (Prescribing Informations Systems, PIS) which is community based. In Scotland, this builds on research using the older SMS1932 survey which did not have PIS linked for dementia ascertainment.

A further strength of this thesis was the ability to learn about, and use data in two different countries (ANZ and Scotland) and become familiar with the practical aspects of the the Five-Safes Framework in two different countries and the variations in permissions and accesses and processes. In the UK, the recent considerations of streamlining the processes and approvals for researchers to use routinely collected data for research highlights what I have experienced, and discovered that there are opportunities to learn from different national systems to progress health data research.

A strength of my thesis is my clinical background for this research. Assessment,

management and care for those living with dementia is a large, and important, part of my speciality, geriatric medicine. I have undertaken memory clinics, and am familiar with key cognitive assessment tools for diagnosing dementia. I also have an understanding of the challenges of the practical clinical side of recording dementia in medical records and how this can be translated to hospital electronic records within the datasets (which is the likely reason for the high proportion of ‘unspecified’ dementia). I am familiar with the clinical aspects that may mean an individual has a diagnosis of dementia but does not have it coded on a discharge letter or death certificate.

When I started this PhD, the use of routinely collected health data as a resource for dementia epidemiology research was growing. A strength is that my PhD was at a time of transition in the growing development of routinely collected healthcare data as a valuable resource for dementia research, and developments in streamlining the governance and ethics processes around access to routinely collected health and social care data for research.

There have been significant changes in the overall structure of healthcare data research in the UK, and also discussion around the standardisation of structures within big data research such as accessing Trusted Research Environments (TREs) also known as National Safe Havens. Part of the valuable learning of this PhD has involved linking and accessing routinely collected data, including different approaches, in different countries (Scotland and ANZ).

Limitations:

A limitation of this thesis is that it only investigates an association between rural and urban living and dementia. It is not designed to a) investigate a causal pathway or b) investigate variables within the rural or urban setting that might be contributing to the association. In Section 7.6 ([Future work](#)) I discuss plans for future research that could overcome some of these limitations.

Overall limitations of using routinely collected health data for ascertaining dementia include: the diagnostic methods used for the diagnosis of dementia, and the ‘diagnostic gap’ (i.e. how many have dementia but do not have it recorded in the datasets I have used). They are linked but distinct issues.

It was difficult to ascertain the accuracy of the dementia diagnosis from different sources, as I did not have access to the information used for confirming the diagnosis of dementia. It is known that those with dementia recorded in routinely collected health datasets have a high positive predictive value for dementia (if dementia recorded they are likely to have it), but this work has been undertaken in other HICs not ANZ.

I had the information on death certificate data only for 2012 in my research in ANZ. I also had information on the hospital discharges and drugs. I could have investigated this further to triangulate how accurate death certificates were in ANZ for recording dementia if they had a diagnosis recorded in one of the other datasets. This was unfortunate, but in part limited by time, needing to be in ANZ to use the secure datalab and also events outside of my control such as Cyclone Cook, closing the secure data lab facilities for some days. It is accepted that death certificates may have patchy reporting of dementia, and it is unclear how this translates between countries. It may also vary between individual medical practitioner certifying the death.

Not everyone with dementia will want, seek or need, a diagnosis, and this is unlikely to be at random, and will be influenced by factors such as age, frailty, individual and family expectations, and access to healthcare services. This can introduce bias when investigating dementia epidemiologically as highlighted in [Chapter 1: Introduction](#). Therefore, assumptions made from results need interpreting with this in mind as potential unknown bias may be introduced unintentionally in a study design.

Whilst a strength of my thesis was using several routinely collected datasets for dementia ascertainment, in both [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#), a limitation was that I was not able to use primary care (General Practice) records, but primary care records are now becoming accessible in TREs, and future research can include these.

Challenges:

Challenges that have common themes across this thesis include accessing routinely collected data in a timely fashion to produce relevant new information to inform healthcare and understanding of a disease, whilst respecting the need for the important steps to ensure the Five-Safes Framework for data are adhered to. It is a difficult balance. Recently in June 2024, the UK had a major data breach within health data as part of a major global IT outage [[700](#), [701](#), [702](#)]. The National Safe Haven in Scotland was not affected. It predominately affected two NHS Trusts in South-East England with substantial impact on clinical care and involvement of National Crime and Cyber Security agencies to recover the data [[700](#), [701](#), [702](#)].

There have been other data breaches in Scotland from hackers including in an NHS Scotland Healthboard in February 2024 [[703](#)]. In the past there has been caution about linkage of data, particularly in England with the *care.data* proposal in 2013, and loss of trust (a key component of the social licence in using data) from the population [[704](#)].

A challenge moving forwards is linking data for the public good and for advancing research in diseases such as dementia in a meaningful way whilst ensuring legal compliance, GDPR and having the public on board with this. Communicating how it is done and the reasons for it, as well as engaging patient and public groups as part of Patient and Public Involvement and Engagement (PPIE) is vital for this to be successful. People in Scotland are broadly supportive of the use of

healthcare data for research purposes [705].

ANZ continue to review how data are used within their IDI, involving different stakeholders, and the public's opinion and trust in Statistics New Zealand to use their data [617, 706, 446]. It is important the PPIE includes recognition of the different ethnicities within a population. I have learnt about the additional role of Māori data sovereignty, and the importance of this in driving forward policy for using data, including data within the IDI for researchers [615]. I discuss this more in Section 7.5 (Implications of my research).

Another challenge is how to define a rural or urban area so that there is uniformity when comparing different countries. I discuss this in more detail in Section 7.5 (Implications of my research) as this challenge is a major implication of my research findings for future work and an area of growing change in recent years.

A final challenge overlaps with the limitation I have outlined: who has symptoms of dementia, but does not have a diagnosis of dementia. I discuss this further in the implications section of this Chapter. However, I give a summary of the challenge here. As discussed in Chapter 3: Using big data for dementia research, people with dementia are often, by the nature of their disease, excluded from national surveys, and undertaking face-to-face epidemiological studies incur significant costs. Therefore, in recent years, data linkage to electronic health records as a means to establish dementia ascertainment has become a growing area of interest to obtain prevalence information on key diseases at a population level. For dementia ascertainment, it can be a valuable method of understanding the disease prevalence when those with dementia would not, due to the nature of the disease, feature traditionally in national surveys of prevalence.

Chapter 1: Introduction highlighted some of the complexities of defining, and diagnosing dementia in a clinical setting. The *pre-clinical* phase of Alzheimer's disease may be years to decades prior to the emergence of clear clinical symptoms

of dementia. Furthermore, not everyone with a *pre-clinical* or *prodromal* phase of Alzheimer's disease will progress to dementia. Some individuals in the *prodromal* phase (Mild Cognitive Impairment) may remain in this phase. A challenge in routinely collected health data is not being clear at what point in a person's life they developed dementia. There may be a coded diagnosis entered in health records, which represents the date they were diagnosed with dementia. This may be influenced by a delay seeking a diagnosis, a delay in referral, or a delay in waiting lists to be seen which combined over several years could give a different age of 'diagnosis.'

Attitudes towards dementia vary within countries and between countries including between different ethnicities within a country [647, 646, 131, 672]. Routinely collected health data may not be representative of all with dementia nor all ethnicities with dementia in a country. Medical recording of dementia in routinely collected health data is predominantly dependent on a clinician recording it in medical records/discharges, and coding departments recording the corresponding ICD code. Some clinicians may code the diagnosis also. Too many diagnoses for a patient may result in the principal code, but not all diagnoses/ICD codes, being recorded in discharges.

ICD coding changes with time, both in the versions in use (Version 9, 10 and now 11) but also internationally (the UK and Europe use a slightly different version to the Australasian modified Version and these are updated/modified in different time points with the Australasian modified Version updated in a timelier fashion) [624]. ICD-11 was endorsed by WHO members in 2019 [21, 21], and operational from 2022, but is not yet in use across all of Scotland - although psychiatry are using it. Secondary healthcare data will only show those with dementia who: a) interact with secondary care, and b) have it recorded on their records. All these aspects clearly highlight both strengths and limitations in ascertaining dementia from routinely collected health data, particularly predominantly secondary care health data.

7.4 Impact of my research

From my research in ANZ ([Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#)) I highlighted four key important next steps: 1) the IDI could be used for dementia ascertainment; 2) the need for ANZ to use country-specific national data to model dementia predictions that are representative of the different ethnicities, 3) the need to expand the datasets used for ascertaining dementia in the IDI including the interRAI dataset, and 4) the need for an urgent more detailed national epidemiological study of dementia prevalence in ANZ and that the IDI could be utilised to expand on this.

One important impact of my research was demonstrating that the nationally collected routine datasets within the Integrated Data Infrastructure (IDI) could be used for dementia ascertainment. To my knowledge, this is the first time the IDI was used for dementia ascertainment. By demonstrating it was possible to use this national resource for dementia ascertainment, further work has since been undertaken by ANZ researchers using both my methods for ascertaining dementia and also separately expanding them with additional datasets which I highlighted in the discussion of my published paper in [Chapter 4](#) would be an important next step. Three high impact papers that have cited my methods in ANZ for dementia ascertainment in the IDI [[652](#), [651](#), [357](#)].

A key outcome of this paper was the call for population-wide country specific data to be used as a national epidemiological study in ANZ, and supporting the further use of the IDI as a valuable resource to achieve this. My study highlighted the need to explore national dementia prevalence specific to ANZ as the cautious estimates in my study differed from estimates using other country data modelled to ANZ. As national repositories of data grow, particularly the excellent example of the IDI in ANZ, which could prove a valuable resource for dementia research. The IDI is not without limitations. However, it has the potential (now already demonstrated by linking the InterRAI data since my publication) of integrating

additional datasets. A future direction could be to integrate research studies as nested studies within the routinely collected data in a similar approach to my research in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) in Scotland to further understand dementia prevalence in ANZ reflecting the multiple ethnicities present and using ANZ data.

Linkage of the IDI with interRAI, along with 3 further additional datasets, has found higher ascertainment of dementia with these additional datasets, and a different time period, linked within the IDI [\[707\]](#). The study found a crude dementia prevalence of 3.8%–4.0% in the >60 years population; and 13.7%–14.4% in the age >80 years population across the four study years (2016–2020) [\[707\]](#). Age-sex standardised dementia prevalence for 2019–2020 was 5.4% for Māori, 6.3% for Pacific Islander, 3.7% for European, and 3.4% for Asian in the >60 years population; and 17.5% for Māori, 22.2% for Pacific Islander, 13.6% for European and 13.5% for Asian in the >80 years population [\[707\]](#).

This supports my summary, in the conclusion of the publication of my ANZ work, that further datasets included into routinely collected data are known to increase yield of ascertaining a disease prevalence. Importantly, their results support mine highlighting the disparity in results between Māori and Pākehā (white New Zealander) which needs further work, and underlines the conclusion that prevalence rates for planning policies need to use country-specific data rather than extrapolate from other countries. Further work is being done in NZ on understanding this complex prevalence between different ethnic groups and how to obtain accurate country specific prevalence rates for improving the lives of those people living with dementia within that country.

Overall, my research in ANZ has been a key building block in using the IDI for dementia research, which has since been further expanded, to investigate in more detail the causes and consequences of dementia using routinely collected health data in the IDI [\[707\]](#). This work is similar to that I proposed in a grant application

in 2016 to Brain Research New Zealand (BRNZ) expanding datasets in the IDI for a prevalence study of dementia in ANZ. Further work had been undertaken investigating anti-dementia drugs within the IDI, and capture/recapture methods for ascertaining dementia in the IDI [707, 620, 708].

7.5 Implications of my research

My research has important implications for future work and development in this field. In this section I highlight and explore five implications from the thesis that are of importance.

7.5.1 What constitutes an urban or rural area?

A key finding of the thesis is the need to clarify and unify the definition of what constitutes a rural or urban area. There is much heterogeneity in this definition, both between countries and across time. The result of this is complexity in fully investigating this variable (rural versus urban living) and its association with subsequent development of dementia. Despite this complexity, my research builds on previous work showing that there are differences between those living in rural and urban areas and their subsequent risk of dementia. This highlights there is a *need* to further define and explore this rural/urban difference in future research. There may be key factors driving the differences which could result in potentially modifiable risk factors for developing dementia.

During my PhD, there have been multiple developments within Europe and internationally in identifying what constitutes a rural or urban area, and how to unify this definition internationally. These changes have taken time to be universally accepted. Initially in 2014, the European Union produced a regional working paper entitled ‘*A harmonised definition of cities and rural areas: the new degree of urbanisation*’ [377, 709] which described the new degree of urbanisation (DE-GURBA) classification. Following this, in 2020 the United Nations endorsed the

Degree of Urbanisation variable as a single global method for classifying what constitutes a city, urban or rural areas in any part of the world. This Degree of Urbanisation variable was created by an international coalition of organisations including the European Commission, Food and Agriculture Organisation of the United Nations (FAO), International Labour Organisation (ILO), the Organisation for Economic Co-operation and Development (OECD), UN-Habitat and the World Bank [375, 376].

A key project progressing the DEGURBA classification is The Global Human Settlement Layer (GHSL) [375]. Their mission is to produce *"global spatial information about the human presence on the planet over time."* They create built-up maps, population density maps and settlement areas which are created through new spatial data mining technologies. Last year in 2023, the European Union produced a Joint Research Centre (JRC) Scientific Informations Systems and Database report entitled "The GHSL Data Package 2023" [710]. This package provides updated *"global spatial information, evidence-based analytics and knowledge describing the human presence on planet Earth"* by providing population level information including classification of areas according to the Degree of Urbanisation (DEGURBA) [377, 710, 375]. The GHSL works in partnership with the Group on Earth Observations (GEO), which is an intergovernmental body and collaboration of expertise [711, 375]. The GEO work programme 2023-2025 has four focuses on key areas: climate change, disaster risk reduction, data, and resilient cities and human settlement [712].

Whilst these international advances in defining a unified classification of what constitutes an urban or rural area are important for future research, using them in medical research to investigate diseases, such as dementia, may take time to implement. Firstly, routinely collected health datasets need to have this information as part of their core dataset. My understanding is, through discussions with data scientists and the eDRIS research coordinator, that this is not yet part of the NSH. This highlights an important implication of my research: what constitutes

the minimum data for a routinely collected health dataset for future research and how can meaningful (and translatable to different time periods and countries) geographical data feature within that?

7.5.2 Geography within routinely collected health data

As highlighted in [Chapter 3: Using big data for dementia research](#), I was involved during my PhD, along with other UK clinical academic geriatricians, in a Delphi process considering what was required for a UK core dataset for geriatric medicine research [11]. This was not specific to routinely collected health data, but geriatric medicine research in general. There is an argument that, as the secondary use of routinely collected health data expands as a resource for healthcare research that consideration of what constitutes a minimum data sets for routinely collected health data for ageing health research is important. It is an area that needs further work to develop a uniform position, particularly encompassing an appropriate, and meaningful, measure for location/geography. Comparison between and within countries is important in understanding the epidemiology of a disease. I would argue that the minimum datasets for routinely collected health data need to be further discussed and that it needs to have a meaningful spatial element at a population and individual level.

The Unique Property Reference Number (UPRN) [379] may offer a solution in the UK for this, or the development of the UN Degree of Urbanisation (DEGUBRA) [376, 709, 713]. It needs to be relatable to the new international definition of urbanisation across the UK and worldwide. For research using routinely collected health data using geographical data, it would be helpful for the REporting of studies Conducted using Observational Routinely-collected data (RECORD) guidelines [433] for researchers to specify what is used to define a geographical area (e.g. quantify population size of a rural or urban areas if comparing different cities/areas). This would improve interpretation of the results and also be beneficial for future systematic reviews comparing differences in dementia between rural and urban areas to allow more heterogeneous studies for meta-analysis.

Whilst country boundaries are less likely to vary with time, much of the ‘within’ country defined geographical areas can vary with time. New postcodes are created as built-up areas expand; council areas/electoral wards can be modified with time depending on the population density. Urbanisation of rural or semi-rural areas also alter the dynamics of rural and urban areas. As discussed in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#), countries differ worldwide in how they define rural and urban areas. There is not one uniform definition, although most classifications include similar key factors such as population density and accessibility. Defining a geographical area over time is not always exact and deciding which geographical scale to use is important to consider.

Within the UK, there are differences in how geographical areas are classified and what geographical variables are available in routinely collected health data. Linking external datasets (such as the SMS1947 survey in [Chapter 6: Results of the SMS1947 dementia data linkage study](#)) to routinely collected health data has the potential to include more geographical data than traditionally available in routinely collected health data, such as the Urban-Rural 2020 Classification System [663]. Balancing privacy with detail is complex. Including more detailed geographical data in routinely collected health data can inadvertently make an individual identifiable if the counts of a disease within a small geographical area allow an individual to be identified. Having geographical data as part of routinely collected health data is important for research in investigating the interplay between prevalence of a disease and location alongside access to services and delivery of care.

The hierarchy of geography across the four UK nations varies (both administratively via Office for National Statistics, and other administrative bodies, as well as health-wise via the NHS) [714]. This include what geographical ‘building blocks’ are used to create these [714]. The developments of the Urban-Rural

2020 Classification System in Scotland is an important development in unifying the definition across Scotland, and with consistency across time, in what constitutes a rural and urban area at a population level whilst protecting confidentiality at an individual level. England and Wales have a rural/urban classification first developed in 2001 from the census and later updated in 2011 [714]. It is constructed from lower layer super output areas and built from the Rural Urban Classification at Output Area (OA) levels in England and Wales [714].

Northern Ireland has recently changed from Output Areas and Super Output Areas (created from building blocks of postcode polygons) to creating two new areas following the 2021 Census. These new geographical areas are Data Zones ($n=3,780$) which nests within the Super Data Zones ($n=850$) [715]. This has created a four tier hierarchical geographical scale in Northern Ireland (Data Zones, Super Data Zones, District Electoral Area, and Local Government District, $n=11$) [715].

In 2005, and again 2015, Northern Ireland Statistics and Research Agency (NISRA) considered how to define rural and urban areas in Northern Ireland with the focus on population size and drive time to nearest population [716]. The 2015 review agreed with the earlier 2006, and NISRA recommended not using terms ‘accessible’ or ‘remote’ to define rural areas as Scotland uses [716]. Instead, NISRA continued to favour descriptions of rural-urban definitions based on individual projects rather than producing a rural-urban classification system that is used by all across Northern Ireland [716].

Additionally, OAs that are used in construction of these variables were introduced at different times across the UK. In Scotland they were introduced following the 1981 Census, but were not introduced across the rest of the UK until the 2001 Census [717, 718, 714]. The lowest geographical unit varies across the UK nations also - only Scotland and Northern Ireland have data zones and introduced in different years: 2004 and 2021 respectively. The purpose of highlighting this is

the variability in geographical scales between four nations of the UK means when investigating rural and urban areas impacting on health they are: 1) constructed in different ways and 2) are not introduced at the same time. Considering this at a larger international scale this can be even more challenging when there is not uniformity within the UK.

Even comparison of research using routinely collected health data investigating dementia, and variation between rural and urban areas across the four nations of the UK produces complexity, not solely in terms of data linkage but through variations in geographical variables such as rural/urban definition/classifications, as well as classifications used for deprivation. It is difficult to compare when there is no universal system in use across the four nations of the UK.

When considering this at an international scale for healthcare research in general (and specifically dementia research considering my research question), there is likely to be a significant lag in time before the approved international definitions of urban and rural areas can be used in a meaningful way in healthcare research using routinely collected health data. The development of the Unique Property Reference Number [379] will provide future potential for investigation of rurality/urbanicity without risk of identification. The integration of the UPRN within routinely collected health datasets [379, 535] is an important first step in linking routinely collected health datasets with geographical variables that are consistent over time, are not identifiable, and allow meaningful research between countries in the UK.

Similarly, there are different variables for deprivation across the four nations of the UK when considering what the impact of deprivation within an area might have on investigating rural or urban living and rate of dementia.

Geography changes with time, including what constitutes an areas in terms of access to services, gentrification of previous poorer urban areas (with the dis-

placement of those who may have been living there), urbanisation of previously rural areas due to the expanding need of housing in the UK and migration to rural and ‘commuter regions’ or working from home changing the dynamics of the traditional use of rural or urban areas. The temporal aspect of a spatial variable is challenging to fully explore in a longitudinal life-course approach which is needed for understanding the role of location (rural versus urban living) and subsequent development of dementia.

Integrating a meaningful geographical variable into routinely collected health datasets is important. The Urban-Rural Classification System in Scotland within routinely collected health data is powerful, but there are still challenges to understand where these areas are in Scotland as currently it is possible to investigate variation between rural and urban but not, within my data, to establish if rural areas in Southern versus Northern Scotland had differences in their association with individuals subsequently developing dementia.

7.5.3 Country-specific data for dementia prevalence

Data need to represent the population to advance dementia research, policy, care and services. Future research needs to reflect the communities and diverse ethnicities within the country populations studied. We know that increasing reports in the UK highlight health outcomes are worse for those from ethnic minorities [407, 405, 402]. Hence data must be inclusive of everyone. Risk factors may differ, and subsequent prevention of dementia, including dementia prediction modelling and precision medicine focused on individual potentially modifiable risk factors, needs to encompass a holistic approach [407, 399, 131, 404, 719]. How ethnicity is defined in routinely collected data is important, and must be inclusive of all - including Indigenous populations and recognising that past approaches controlling how someone self-identifies, or indeed recognising multi-ethnic populations, are not inclusive, and can be harmful.

Dementia research is predominantly undertaken in High Income Countries (HICs)

with largely people of white ethnicity. This does not mean the results are translatable to all, and in different country contexts. Using routinely collected health data can be a valuable resource reflecting the health of the nation it serves. There are caveats to this. Firstly, variations between HICs (more likely to be using routinely collected health data) versus L-MICs, and secondly differences in how healthcare is funded (privately or nationally and free) are important to appreciate. As the drive to use routinely collected health data in HICs increases, this could create widening gaps in dementia research between HICs and L-MICs, and even within HICs who is captured within the datasets depending on 1) how healthcare is delivered, and 2) how important demographics such as ethnicity are captured.

A drive to link data and re-purpose routinely collected health data for secondary use in healthcare research provides a unique opportunity to develop systems to capture more accurate country specific estimates of dementia incidence and prevalence. Five important implications from my thesis in relation to the need to use country-specific data for dementia prevalence include:

1. Re-purposing routinely collected health data for research
2. Seeking a diagnosis of dementia (drivers and barriers)
3. Variations in dementia rates
4. Nested research studies within routinely collected health data
5. Differences between HIC and L-MICs

I now discuss these as important implications from my research.

Re-purposing routinely collected data for research

Whilst initially established to provide healthcare and care for patients (through the delivery of healthcare services and monitoring of this), routinely collected health data can be valuable for other reasons. There is a changing direction to secondarily re-purpose routinely collected health data for healthcare research,

including dementia research. This thesis is an example of this with the addition of a ‘*nested*’ further study (the SMS1947 historical survey) with routinely collected health data. Future implications of this include how to further integrate routinely collected health data not just for health services and delivery, but also healthcare research. They all inform one another.

Integrated systems and reduced delays in undertaking research can ultimately help people living with a disease sooner or inform service delivery and health policy. The question is how to maximise re-purposing routinely collected health data, including integration of additional research datasets, for the public’s benefit, and an ageing UK population in which understanding dementia is important for the individual and society, whilst also respecting the constraints and appropriate procedures to ensure the Five-Safes Framework are upheld.

Overhaul of procedures, streamlining and TREs have been helpfully proposed by Torabi et al. 2023 [498]. In Scotland, this has begun with Research Data Scotland [543]. In the UK, the outcome of The Sudlow Review [537] on the future direction, and use, of health data, will be helpful for future health data research. Within ageing health data research, a collaboration I am involved with the Ageing Data Research Collaborative (ADRC), and the European Geriatric Medicine Special Interest Group (EuGMS SIG) on health data research are important for driving this forwards. Engaging multi-disciplinary collaborations of stakeholders (including clinicians) will help shape the future direction of using routinely collected health data for ageing research in the UK and across Europe. Future machine learning or Artificial Intelligence (AI) could be important in understanding free text within routinely collected health data for dementia ascertainment and possible risk factors that can have valuable information that is not coded and therefore often missed [720, 477, 721, 722].

Whilst the re-purposing of routinely collected health data for a secondary use in healthcare research grows, there are limitations in using routinely collected

health data that is not designed specifically for research, and also with a tendency to focus on secondary healthcare datasets. Being able to *nest*, or integrate, a research study/survey within routinely collected health data can be extremely valuable in overcoming the first point [179, 501]. Data within routinely collected health datasets are only as good as what is inputted in the first place from clinical encounter to coded information. A researcher has no control over this when using routinely collected health data (with/without research studies linked to it) when it is being analysed as a secondary outcome in research. To improve this, close working between many groups including clinicians at the front line using the datasets with the patients, medical coders and researchers is important.

The Health Data Research UK (HDRUK) provides a drive for big data investigating complex disease. Indeed they have created new PhD studentships around research specifically for big data and complex disease in the UK [538]. This includes Cardiovascular disease (CVD) and cancer but does not appear to include dementia. Dementia would be a further important complex disease for future work.

Scotland is not large and has one national healthcare service and the the unique identifier, the CHI-number which allows data linkage. Scotland's unique identifier is not like the Nordic countries, which are world leading in data linkage due to their unique identifier across all administrative and health databases. However, Scotland is a similar population size to ANZ which also has a unique identifier the National Health Index (NHI), and has a nationally linked resource for routinely collected data, the Integrated Data Infrastructure (IDI).

In Scotland there are national datasets for some neurological diseases (Motor Neurone Disease - Scottish MND Register [723], and Multiple Sclerosis - Scottish Multiple Sclerosis Register [724]) that could be linked with early onset dementia. I have previously been involved in linking data in the diabetes register using electronic health records [8], another important dataset. A recent study pub-

lished this year (2024) investigates The Scottish Medical Imaging (SMI) Archive, examining 57.3 Million radiology studies linked at an individual level to the individual's electronic medical records [725]. These are some examples of linkage within Scotland that could be further enhanced through the ability to link data with the CHI-number. There is real potential to create a routinely collected data driven dementia research ready national data linkage resource in Scotland with linked data from different research studies to routinely collected health data

There are also national drives for data linkage which were mentioned in [Chapter 3: Using big data for dementia research](#) such as Data Loch in South-east Scotland [536, 552]. It is important that data linkage includes not only healthcare data (both primary and secondary care), but also data from social care, e.g. those resident in care homes or receiving care services in their own homes. These are important groups in society that must not be further isolated from data linkage to improve understanding of dementia [565, 497, 564, 562].

The NHS Research Scotland (NRS) Neuroprogressive and Dementia Network (NDN) connects those working in dementia research across Scotland [726, 727]. Importantly, there is a strong focus on involving lived experience of dementia and neuroprogressive diseases in the goals of the network. The network's strategy for 2022-2027 is holistic and driven to improve understanding of dementia in Scotland with a mission to *"Deliver cutting edge clinical research in neuroprogressive conditions and dementia across Scotland"* [726].

In the 28 commitments of the NRS NDN 2022-2027 strategy there is no specific mention of, or use of the words 'routinely collected health data', or synonyms, in harnessing the power of routinely collected health and social care data across Scotland, or the unique opportunity that Scotland has for data linkage between research studies and routinely collected data because of the CHI-number [726]. There is real potential to synergistically work with the advancements in the National Safe Haven (NSH) in Scotland, and routinely collected health and social

care data, to explore an additional layer for dementia research, and ‘nested’ research studies with routinely collected health data in Scotland.

This could be powerful in combining the HDRUK data drive for research, the Scottish Government Dementia Strategy and the Scottish Government’s Data Strategy for Health and Social Care [547]. The proposal to investigate the use of the new CHI for not just healthcare but also social care in Scotland could reduce, or dramatically improve data linkage in Scotland [541]. A Scotland-wide directory of research studies which have been to date linked to routinely collected health and social care data for dementia research would be an important first step in exploring the data linkage projects in one place. ANZ has a similar summary for different diseases using their national linked data resource, the IDI.

From my research in ANZ, I believe there is real potential to do similar as has been done with cardiovascular diseases and the Vascular Informatics using Epidemiology and the Web (VIEW) network in ANZ but for dementia research creating a dementia health and social care data network [728, 604]. There is also potential to link additional datasets to the IDI as has been demonstrated with the interRAI [630] dataset, and I explored the linkage of memory clinic data as discussed in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#).

In summary, an implication of my thesis is the importance of comparing and exploring how different countries re-purpose routinely collected health data for research, with a focus on ANZ and Scotland in this thesis. There is real potential to investigate dementia further, and variation across Scotland, to inform the disease process, health services and delivery of care. It is important that re-purposing routinely collected health data for dementia research involves multiple stakeholders, and critically patient/public involvement. It is also essential that health and social care are linked in the future for driving forwards dementia research.

Seeking a diagnosis of dementia

When comparing countries, not all have the same national drive or focus on dementia alongside variations in cultural approaches to wanting a diagnosis of dementia. National strategies for dementia diagnoses, and management, may provide drivers for individuals to seek a diagnosis. Conversely, cultural or ethnic differences in having a dementia diagnosis might prove barriers alongside drivers. There are numerous reports on factors influencing individuals seeking a diagnosis of dementia [729, 525, 672, 647, 730, 404, 131].

I discussed some of the recent research highlighting drivers and barriers for those seeking a diagnosis of dementia in ANZ in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#). This is important to consider in the context of ‘diagnostic gap’ in dementia diagnosis and how this might influence dementia ascertainment in my research in this chapter. More recently, the impact of the Coronavirus 2019 (COVID-19) pandemic has suggested additional changes in the rates of dementia diagnosis within the UK with decline of 5.4% in the estimated Dementia Diagnosis Rate (DDR) from March 2020 to February 2023 in England [731]. This is important for future studies using routinely collected health data in the UK to investigate dementia incidence and prevalence.

Within HICs and L-MICs there may be barriers or drivers to seek a dementia diagnosis [672, 128, 671]. Even within HICs, there can be recognised under-diagnosis as a paper this year highlights. This recent study in Australia highlighted under-diagnosis of dementia in GP records [732]. Out of 2,525 people with a clinically validated free text diagnosis of dementia, 372 (14.73%) people had no coded dementia diagnosis [732].

Under-diagnosis in routinely collected health data can be across all datasets for dementia highlighting: 1) a need (as mentioned in my publication, Section 3.2) of

engaging those at the frontline with patients to increase the accuracy of clinical coding so that routinely collected health data is research ready for incidence and prevalence studies to improve *quality* of dementia recording; 2) to increase the yield of dementia ascertainment within routinely collected health data by increasing the number of datasets as much as possible. There is a need to include as many datasets as possible, to improve the *quantity* of dementia recording.

Variations in dementia rates

It is recognised that whilst worldwide prevalence of dementia is rising [92], that there are variations in dementia rates as highlighted in [Chapter 1: Introduction](#). Over the last two decades, some studies suggested that age-specific dementia incidence and prevalence in HICs (Europe and North America) may be declining [90, 89, 733] versus L-MICs where it was rising [92, 80, 81, 82, 22]. However, there are mixed opinions in Europe and North America regarding future predictions of dementia prevalence as it has been suggested whilst dementia incidence has decreased, prevalence has not [90, 84, 85, 89, 88, 87, 86].

One implication of my thesis highlights is the need to use country-specific data to inform country specific dementia prevalence, while learning from international variation in dementia prevalence. Whilst there may be untapped resources of routinely collected health data to investigate national dementia prevalence, as I found in my research in [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#), it is also important to consider other factors that might be driving variations in dementia rates from a data perspective that is not due to under-diagnosis of dementia. This also might drive different directions of association between rural and urban areas, and subsequent risk of developing dementia.

Population dynamics and movement changes over time, perhaps more so now than in previous centuries when migration was most likely to be due to significant world events, such as World Wars, famine or disasters, than movement due

to work or lifestyle choices. Scotland’s population is growing, but only slightly, due to immigration rather than increasing birth rate [734]. Since 2006, for those migrating to Scotland, Edinburgh and Glasgow have remained the most popular destination for overseas migrants [735]. These are large urban cities. The SMS1947 survey used in the research in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) is a historical cohort, and predominantly a white ethnic group. One implication of my research is that this may not be generalisable to the current demographics of Scotland, and also to other HICs with a multi-ethnic population including Indigenous peoples. A relevant future implication of my work is recognising that changes in migration to Scotland, and different ethnic groups (alongside how ethnicity is defined in country official statistics) may require important planning to ensure holistic inclusion of everyone living with dementia within Scotland.

Age-standardisation allows the comparison of dementia rates between populations with different age structures. This is particularly important for diseases such as dementia that increase with age. In Scotland, age standardisation of deaths per 100,000 is reported [736]. To report this, Scotland uses the European standard population 2013 for age standardisation for dementia deaths in Scotland [736].

In ANZ they use standardisation to the World Health Organisation’s standardisation, unless Māori are the focus of the analysis, and non-Māori are the comparator group [737]. As the ANZ Ministry of Health outlines in their report: *Kaupapa Māori research acknowledges the right of Māori to good health and, therefore, advocates centralising Māori in a research setting. The use of the non-Māori group as a comparator for analysing Māori health data is favoured.* [737]. ANZ uses the ‘the direct method of age standardisation’, which is the calculation of an overall standardised rate obtained by applying age-specific rates of a population to a standard population [737]. ANZ Ministry of Health highlights that age-standardised rates are important for populations with different age structures: e.g., Māori (on average younger), and non-Māori populations [737].

European countries do not necessarily reflect other HICs, particularly countries with Indigenous populations. There is a need to develop age, sex and ethnicity standardisation relevant to the populations the data represents. Research in 2007 discussed an Indigenous standard for age standardisation in ANZ for Māori and non-Māori populations [738]. In Australia, it is recommended that for Indigenous Australians the direct standardisation for age standardisation is used in administrative data collections [738, 739]. I was delighted to recently connect with a clinical researcher (Dr Antonia Clarke) at Monash University, Australia, who is embarking on a PhD investigating this important area of age standardisation that is relevant for Indigenous populations.

The aim of Dr Clarke’s PhD is investigating dementia prevalence according to rurality for Indigenous peoples across HICs. The aim is to age-standardise the prevalence ratio between the Indigenous and non-Indigenous populations within each country. As highlighted in my publication in Chapter 4 (Section 4.3), it is vital to recognise the importance of fully including a multi-ethnic approach to dementia prevalence and future predictions. As highlighted in the discussion in Chapter 4, research is suggesting an earlier onset (<65 years) for late onset dementia in ANZ for some ethnic populations [740, 131, 719]. It is essential to understand these variations. By doing this, modifiable risk factors can be appropriately addressed at the most appropriate point in a person’s life-course.

7.5.4 Accessing routinely collected data

With the growing drive to re-purpose routinely collected health data for health-care research, this means health informatics has become an expanding field since my PhD began. Alongside the increased drive to access, link and use routinely collected health data for research, there are increased drives to investigate how to link datasets within and across countries, whilst also protecting privacy, conforming to GDPR and working with different data custodians/controllers responsible for the datasets.

Having worked in two different Trusted Research Environments (TREs) in Scotland and ANZ for this thesis, I have greater understanding of the variations in using different secure data environments, as well as differing approaches to accessing and linking routinely collected data. As highlighted in [Chapter 3: Using big data for dementia research](#), even within the UK there is a wide variation in TREs, alongside variation in the processes for approving a researcher to access and use the data [498], and this is an area recognised to need greater harmony within the UK. This also has important implications for working between countries, particularly in view of my thesis findings, and need for further research to understand rural versus urban differences impacting on developing dementia in the future. Having a uniform definition on what is classed as a rural or urban area is needed, and also similar methodologies, and reporting, (for instance in line with RECORD [433] guidance particularly in HIC where predominantly routinely collected data were used) to allow comparisons in meta-analysis.

Timely access to routinely collected health data and linked datasets is important for a healthcare researcher to maximise the time for investigating, and analysing the research question, to further understanding in a disease for those living with a disease. It can be complex adhering to the Five-Safes Framework in routinely collected data: to both protect the confidentiality of data, whilst balancing timely access to data for research. This can be complex simply within one country, and even more so between countries, when linking different datasets with different data controllers. Processes are in place which are mandatory, but at times can hinder the timely access to the data for the researcher.

Understanding how the processes can incorporate reducing the time for researchers to access the data whilst following the Five-Safes Framework can be challenging. This will also vary between countries, as I have experienced in my PhD, depending on the infrastructure of how the routinely collected data is linked and format made available for the researcher to use. Additionally, the systems are

not designed for cross-country analysis through data linkage in routinely collected health data due to many factors including legal frameworks.

Scotland and ANZ take a different approach to data linkage. In [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) and [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#) I have outlined the specifics of each process in each of these countries to access data for my research. In summary, ANZ (through Statistics NZ) makes the data accessible in a more timely fashion for the researcher (estimated around six weeks when I undertook my research) as the researcher is authorised to access and view certain datasets within the secure TRE (datalab) that they need to create their research study. The researcher is then responsible for creating their denominator (the database and study population they are investigating) and numerator (those with the disease).

Conversely in Scotland, the researcher is authorised access to only view their already created denominator (*database* population). This has been created by data analysis working for eDRIS. The researcher then creates their study population and numerator (those with the disease). The data are more curated for the researcher in Scotland but is dependent on a sufficient number of, and time of, data analysts to do this important first step whilst they also will be doing similar for other researchers. There are positives and negatives of both options and ultimately depends on the processes that have been agreed within the Five-Safes Framework in different locations and how the public are willing for their data to be used.

In ANZ, Statistics NZ commissioned Opus International Consultants Ltd on their behalf in 2015 to examine public attitudes to the IDI and use of personal data for research. [\[494, 631, 618\]](#) It involved a multi-method qualitative approach which included online discussion, workshops and in-depth narrative interviews. There is no agreed definition of what ‘social licence’ means. A key component, essential

for using routinely collected data, is that the public has trust in those overseeing the data (data custodians).

Overall, the findings from Statistics NZ report suggested that the public were happy with their data to be used if it could benefit people and could remain de-identified, and not shared with external organisations [494, 631, 618]. However, there are suggestions that Statistics NZ’s social licence for the IDI could be described as ‘tenuous’ as a proportion of the public has no view (positive or negative) towards the use of their data [616]. Further work is being done to update this, and importantly ongoing assurances of including Māori Data Sovereignty within the IDI [618, 706]. In Scotland, as outlined in [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#), access to data is controlled through the Health and Social Care Public Privacy Benefit Panel (HSC-PPBP) [452].

In addition to the timely access of linked data within different countries, there is also the challenge as highlighted above of data linkage between countries. Interesting previous developments in navigating this complex field of data linkage include projects such as DataSHIELD and OBiBa [741, 742, 743]. These projects started 10 years ago in a *"joint development of an open-source analytics platform that enables and simplifies flexible but efficient federated analysis"* [744, 745, 746, 741, 743]. DataSHIELD has been proposed as a platform that could be used to navigate this as research studies stay on a server at each institution responsible for the data (data controllers) and is then linked to an Opal database (data integration system used by epidemiological studies).

A pan-European cohort built on existing and new large scale population cohorts for COVID-19 was developed called the ORCHESTRA project [747]. It is funded by the European Commission. Separately, the European Health Data Space Initiative [157, 748] is aimed at interconnecting health data silos for *"versatile secondary use of health data"* and benefits from the ability of High Performance Computing (HPC) by data analysts. Within Scotland, the Edinburgh

International Data Facility (EIDF) includes new developments such as DataLoch previously mentioned in [Chapter 3: Using big data for dementia research](#) which is a collaboration of South-East Scotland region's Local Authority, NHS Health Boards, University of Edinburgh, and Health and Social care [536].

The implications of my thesis are that these could be very valuable collaborations of data developments of future importance for my research question. The COVID-19 pandemic during my PhD definitely accelerated an understanding of the importance of routinely collected health data to inform knowledge of a disease and healthcare decisions. It was a unique example of real-world rapid sharing of data improving lives through research by highlighting how timely access to real-world observational linked data could be vital for preventing transmission of COVID-19.

For instance, the ZOE COVID-19 Symptom study app established by Prof Tim Spector at UCL consisted of a free app that could be downloaded by members of the public who completed risk factors and symptoms they were experiencing [749]. One example was the benefit of timely access to data for improving lives such as asking participants about symptoms of a loss/reduced smell (anosmia/hyposomia) [749]. There were case reports and media reports that this might be important but national guidance did not change as there was no data. This real-time use of population data clearly showed how important this was as a marker of COVID-19 with evidence that loss of smell was the most predictive symptom of having a positive COVID-19 test [749]. Overwhelmingly, this example of using health data in a timely fashion without a doubt saved lives and became included in national guidance [749]. During a pandemic, rapidly learning from one another in different countries about a disease and linking datasets became a priority.

The pan-European cohort of linked datasets in COVID-19 through the ORCHESTRA Project [747] shows the potential to overcome silo working in health data,

connecting European cohorts, to understand a disease and improve individuals lives. In terms of dementia research, there is always potential to learn from this and try to navigate improving data linkage within countries and between countries. Alzheimer Europe has outlined the need for data sharing for dementia research in their report, and are expanding the practical ways of doing this across Europe [750]. Ultimately, routinely collected health data re-purposed for research must be used in the public benefit as it is the public's anonymised data, but it is important to consistently review if there are more timely ways of accessing data for researchers whilst adhering to the Five-Safes Framework.

7.5.5 Exploring rural vs. urban risk factors for dementia

There are many differences between rural and urban areas. As discussed in [Chapter 1: Introduction](#), there are geographical determinants of ageing, cognition and brain health that have been investigated. Some of these variables could be potentially linked more with living in a rural or urban area. This poses the question, is rural versus urban living simply representing other variables that those environments encompass? For instance, living in a rural area of a HIC may be associated with less exposure to air pollution, or a more active lifestyle, or increased exposure to green space, all of which have been associated with lower dementia risk and increased well-being. Additionally, if people stay where their families have lived, there may also be genetic aspects to that association.

Rural and urban living could be construed as simply a proxy for many of these potentially modifiable risk factors [97]. As humans we are intricately connected with the environment we live in. It has a role in our health and well-being. We directly interact with our surrounding environment. But, as research has proposed (and highlighted in [Chapter 1: Introduction](#)), the environment also directly, or indirectly, interacts with us. The interaction is complex in terms of our individual health (physical, mental and psychological), but also in relation to our individual gene-environment (specifically epigenome) interaction. As mentioned in Section 1.5.2 ([Location: geography and brain health](#)), the epigenome (unlike

the genome) is environmentally modifiable.

A recent review [269] exploring the rural-urban divide and epigenetic research discussed the rise of non-communicable diseases (NCDs) in L-MIC. Within that it discussed the link between the degree of urbanisation (urbanicity) and NCDs supporting the idea that environmental factors may be critical in the aetiology of NCDs. That review paper proposes *"the possibility exists that differences in epigenetic regulation between urban and rural-dwelling individuals could be the basis by which the differences in environmental exposure shift disease prevalence with urbanicity"* [269].

Multiple potential drivers of epigenetic changes have been identified that could explain some of the rural/urban differences, e.g. air pollution [327]; the gut microbiome may play a role in subsequent risk of developing dementia [326], and future work should explore these potentially interacting influences, and when in the life-course they may have most impact on different outcomes.

The implications of my thesis means there are many questions that need further explored. Does living in a rural or urban area have risks for dementia and do these vary between HIC and L-MIC because the environment differs between these areas in HIC and L-MICs? Does living in a rural or urban area influence dementia risk through environmental factors modifying an individual's epigenetics, and subsequent risk of developing dementia? How is this affected by the 'accumulation of risk' or 'critical/sensitive' period epidemiological life-course perspective? If the environmental factors of living in a rural or urban area are involved in influencing our epigenome, then at what point in a person's life could this be most affected or influenced by where you live or is it accumulative?

Other considerations to explain variations in rural and urban living and dementia diagnoses include influences such as 'push' and 'pull' factors for individuals moving to or from rural/urban environments (e.g., moving to access care homes,

retiring to rural environments). Are these potential factors influencing a risk of bias in interpreting rural/urban variations of dementia diagnoses if individuals with dementia have more care needs requiring a move to a more urban setting, or individuals with less health/social care needs are more freely able to change from urban/rural living environment, or variations in rural/urban environments and dementia diagnoses reflect socio-economic factors (as highlighted in research between HICs and L-MICs). These are important implications of furthering understanding of the role of living in a rural or urban area and subsequent risk of developing dementia and areas for future work.

The 2024 Lancet Commission on Dementia Prevention suggested that dementia could be reduced in approximately 45% of cases through twelve Population Attributable Fractions (PAFs) [97]. Additionally, in 2024, a recent systematic review by Mukadam et al. showed changes in prevalence and incidence of dementia amongst twenty-seven papers with cohorts, and considered PAFs contributing to this [384]. None looked at living in a rural or urban area.

In dementia risk prediction modelling, a recent systematic review [253] highlighted one paper that used rural living in the prediction model. This paper (2018) [302] used data from the Mexico 10/66 cohort study and developed a dementia prediction model that was deemed *fair* in its discriminative performance (Area under the curve, AUC, range: 0.72-0.75) by the systematic review authors and used variables including rurality (along with neuropsychiatric symptoms, illiteracy, mild cognitive impairment and diabetes). The systematic review [253] also highlights that different risk factors may require testing for L-MICs which may be less applicable to high-income settings, but do not specify differences between rural or urban areas. Given the only dementia prediction model to use rurality as a risk factor in their systematic review was undertaken in a L-MIC using 10/66 protocol data in Mexico, I think my thesis supports the need to consider risk factors differences between HIC and L-MIC between rural and urban settings.

In summary, the implications of my thesis in terms of exploring rural versus urban risk factors are important. As Population Attributable Fractions (PAFs) are highlighted as population level risk factors for dementia prevention and dementia prediction models are increasing in number (although with concerns on how translatable and valid they are) and incidence and prevalence of dementia is varying between HIC and LMIC then the implications of my thesis are important in terms of considering the place, or environment, the individual lives in (HIC versus L-MIC, and rural versus urban) and how these may influence individual and population level risk factors for dementia prevention.

7.6 Future work

In the prior section [Implications of my research](#) I highlighted and discussed important implications of my findings for future research in this area. I have also discussed additional future work in the individual research projects of my thesis in [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#), [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#) and [Chapter 6: Results of the SMS1947 dementia data linkage study](#). In this section I focus on specific future work I have already started the process of undertaking to expand and explore the results from the thesis. Following this, I highlight some additional thoughts from drawing together the overall findings from my thesis for future work.

I plan to update the systematic review ([Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#)) for two reasons: 1) further increased research looking at geographical variation of dementia between rural and urban living since the last search, and 2) the development of recent advancements (and recommendations) in geographical search filters for investigating geographical studies in systematic reviews. This will provide a definitive synthesis of the data on the role of rural versus urban living and subsequent dementia, and

explore the role of HIC and L-MIC, and data collection type.

I am building on my results from [Chapter 6: Results of the SMS1947 dementia data linkage study](#) through new permission to obtain geographical data in additional datasets. This will add to the prior work which could not include geographical data for all of the *database* population because there was no access to geographical data in the NRS Mortality or PIS datasets. In December 2023, I requested an amendment to access geographical data in the additional datasets within the NSH which are part of my research project. This amendment was approved in early 2024 and at the end of April 2024, the Urban-Rural 8-fold 2003-2004 variable was added to the missing datasets (NRS Mortality and Prescribing Information System). This will allow my denominator of those with geographical data within the routinely collected CHI-linked SMS 1947 to be larger. I will repeat these results to assess if my findings from [Chapter 6: Results of the SMS1947 dementia data linkage study](#) are similar. Further work following this will investigate different time points in a person life in relation to their geographical location (rural and urban living) and future risk of dementia.

As part of the future work, it aims to capture location (rural vs. urban living) at different points in life. This will be undertaken now geographical data in different datasets (NRS Mortality and the prescribing database, PIS) is available within the National Safe Haven for my study. As the data starts from mid life (when the CHI-number started in the 1980s, the study population were in their early 40s) I will not be able to investigate childhood versus adulthood rural/urban living as part of my study design. As discussed in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) this would be important additional work to investigate critical, sensitive or accumulative risk of rural/urban living and subsequent development of dementia. It would also be important in considering how those rural or urban areas may have changed over time in terms of what constitutes that area.

7.7 Overall conclusions

In conclusion, my thesis investigated the variation of dementia rates/risk between rural and urban areas using routinely collected health datasets to answer my research question: *‘Does dementia vary between rural and urban areas and how can we use routinely collected health data to answer this question?’*. In answering my research question, my thesis found that there are complex relationships between rural and urban living and dementia diagnosis, which are important to consider, and that routinely collected health data are useful in answering this question, but need to be used by researcher who understand the complexity of the data, and the clinical syndrome of dementia. As demonstrated, there are challenges with the definition/uniformity of what constitutes a rural or urban area, and with geocoded data in routinely collected health data.

Key future work in Scotland includes creating the ability to use detailed data from Scotland to estimate dementia prevalence at a population level. As Scotland’s NSH develops, health and social care is linked and the language for coding diagnoses can be used by different systems (e.g. SNOMED CT [475, 548]) facilitating the analysis of primary and secondary care, then Scotland has a real opportunity to use the CHI-number at a population level to gather meaningful dementia population prevalence. As ANZ has a nationally collated resource in the form of the IDI, there is potential to create a similar approach within a NSH in Scotland. The ability to link datasets, particularly research datasets, to this NSH is powerful. In addition to facilitating research to understand dementia further, it would also mean that healthcare policy and health services could be informed by the data within a country.

Overall, the thesis highlights that to progress understanding of the role rural or urban living has on dementia risk, further work is required to: 1) clarify definitions on what constitutes rural or urban areas which are also consistent over time and comparable at an international scale between different countries; 2) link

research datasets to national routinely collected health data to provide a longitudinal life-course research perspective; 3) understand differences between rural and urban areas and dementia rates in HIC/L-MICs; and 4) obtain country-specific estimates using country specific data to ensure representation of the ethnicities within the population.

The interaction between an individual and their environment is complex. There are both direct and indirect interactions between the individual and the environment. This may also change over our life course. The environment may also impact on our epigenome via gene-environment interaction thereby influencing environmental risk factors and our subsequent health.

This complex interplay needs to be investigated further both within countries and between countries. This includes between HIC and L-MICs given the variations in prediction in dementia prevalence forecasting and because the constructs between urban and rural living might be different. Having a uniform definition of urban and rural areas that is internationally accepted and translates to being usable in a medical research context is important. Where we live can affect our health: both at a place in time (spatial), and over time (temporal). Growth in data linkage, and routinely collected health data, are invaluable for exploring population level or large size datasets. Geographical data adds additional level of complexity in ensuring individuals are not identifiable. Future direction for data linkage at a national scale is important so countries can produce meaningful estimates based on the data of the population they represent and the role of rural or urban living can be fully investigated.

Dementia prediction modelling still requires validation and consensus, but as this develops alongside the realisation of how important potentially modifiable risk factors are, there remains a sparsity of research investigating rural and urban differences in dementia rates and risk that may be important for the future. The recent move towards uniform classification of what constitutes urbanisation

internationally provides a unique opportunity to capitalise on this and begin to include a marker for urban or rural living in dementia studies.

Key implications from this thesis include: 1) how urban and rural living is defined and how this definition compares between countries as it is often very heterogeneous, 2) the need to understand the difference in variation in both HICs and L-MICs separately, and 3) the need to capture life-time geographical location, particularly geographical location at particular key points in a person's life such as childhood and mid-life.

Dementia does vary between rural and urban areas, and routinely collected health data are extremely useful to explore this, but require good methodological knowledge. Future work exploring underlying mechanisms will help to inform efforts to reduce the incidence and impact of this devastating disease.

Appendix 1

This Appendix relates to [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#). The systematic review was registered on PROSPERO [342]. I include a pdf copy of the registered protocol on PROSPERO:

Geographical variation of dementia: systematic review with meta-analysis

Citation

Katherine Walesby, Candida Fenton, John Starr, Susan Shenkin, Jennifer Burton, Tim Wilkinson, Luke McGillicuddy, Duncan Gray, Tom Russ. Geographical variation of dementia: systematic review with meta-analysis. PROSPERO 2016 CRD42016050323 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42016050323

Review question

To identify geographical variation and geography as a risk factor for developing or protecting against dementia.
To describe the quality of evidence and the strength of geographical location as a predictor of dementia.

Searches

The following databases will be searched:

ASSIA - Applied Social Science Index;

Ovid MEDLINE;

Ovide EMBASE;

EBSCOHost CINAHL;

GeoRef;

Global Health;

LILACS;

PsycINFO;

COPAC;

SciELO;

ProQuest Dissertations and Theses;

PapersFirst;

ProceedingsFirst.

Types of study to be included

Including: Cross-sectional and longitudinal studies of any length. Studies offering a comparison of dementia prevalence or incidence between two or more different sites, any geographical scale. Excluding: Studies with no comparison between two or more different geographical sites. Studies using external comparison groups (e.g. European Community Concerted Action on the Epidemiology and Prevention of Dementia/European collaboration on Dementia papers).

Condition or domain being studied

Dementia.

Participants/population

Including: People with all causes of dementia except those where dementia is a secondary feature of a chronic disease.

Excluding: Dementia due to chronic disease (alcohol or traumatic brain injury, Parkinson's disease, Huntington's disease and Creutzfeldt Jakob diseases, either sporadic or variant). Studies focusing purely on young onset dementia (less than 60 years old).

Intervention(s), exposure(s)

The exposure to be reviewed is geographical location.

Comparator(s)/control

Inclusion: Any studies offering a comparison of dementia prevalence or incidence in people between two or more different geographical sites, at any geographical scale.

Exclusion: Studies where there is no spatial variable for the population (e.g. comparing different ethnic groups or investigating concentrations of pollutants in water).

Context

Studies looking at dementia prevalence or incidence between two or more geographical locations.

Main outcome(s)

Geographical locations of developing dementia and their associated risk ratios (or other measure of predictive power).

Prevalence or incidence of dementia in the two (or more) areas studied - including number of cases and non-cases.

The scale of comparison or areas that were compared - including place/country of study. Diagnostic criteria used for dementia diagnosis.

Measures of effect

The timing is geographical location at any time point(s) during life course. Timing (i.e. age) will be related to exposure (geographical location) to infer possible epochs of risk.

Additional outcome(s)

None.

Measures of effect

Not applicable.

Data extraction (selection and coding)

Titles and abstracts will be reviewed by two independent reviewers using the inclusion and exclusion criteria to identify studies and obtain relevant full-text versions. The search will be recorded with a PRISMA diagram. Two reviewers will determine studies for inclusion. If there are any discrepancies in the initial title and abstract screening assessment then both will be included for full text review.

Any discrepancies in the assessment of full text articles will be subsequently resolved by a third independent reviewer.

Data extraction will be performed using a specific template for this review.

Risk of bias (quality) assessment

Risk of bias assessment will be performed on each study using an appropriate tool. Additional aspects of quality assessment not covered by risk of bias assessment alone, will be extracted from other appropriate tools used for

observational and non-randomised study quality assessment.

Strategy for data synthesis

We will provide a narrative synthesis of the included studies and their findings.

This will be structured around the type of study, country or geographical location studies and specific information on the participants such as demographics, the number of participants, and the participants' ages and sexes. Additional information relevant to dementia risk will be included, such as smoking status, the history of cardiovascular events, and the history of cerebrovascular events. Lastly, the level of cognition and cognitive scores (and tests used) will be included in the narrative.

A quantitative synthesis will be used if the included studies are sufficiently homogenous. Where studies have used the similar methodologies, we will pool the results using random-effects meta-analysis.

Heterogeneity between the studies in effect measures will be assessed.

Analysis of subgroups or subsets

If the necessary data are available, subgroup analyses will be undertaken.

Depending on the included studies, the analysis will be decided using the most appropriate subgroups which will be determined by both reviewers (e.g. rural versus urban, latitude, country specific etc.).

Contact details for further information

Katherine E. Walesby

katherine.walesby@ed.ac.uk

Organisational affiliation of the review

None

Review team members and their organisational affiliations [1 change]

Dr Katherine Walesby. Alzheimer Scotland Dementia Research Centre and Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh

Ms Candida Fenton. MRC/CSO Social and Public Health Sciences Unit, University of Glasgow

Professor John Starr. Alzheimer Scotland Dementia Research Centre, University of Edinburgh

Dr Susan Shenkin. Geriatric Medicine, Centre for Clinical Brain Sciences, University of Edinburgh

Dr Jennifer Burton. Academic Geriatric Medicine, Institute of Cardiovascular and Medical Sciences, Glasgow Royal Infirmary

Dr Tim Wilkinson. University of Edinburgh

Dr Luke McGillicuddy. NHS Lothian

Dr Duncan Gray. NHS Lothian

Dr Tom Russ. Alzheimer Scotland Dementia Research Centre, University of Edinburgh

Type and method of review

Epidemiologic, Meta-analysis, Systematic review

Anticipated or actual start date

11 November 2016

Anticipated completion date [1 change]

31 December 2021

Funding sources/sponsors

None

Conflicts of interest

None known

Language

English

Country

Scotland

Published protocol

http://www.crd.york.ac.uk/PROSPEROFILES/50323_PROTOCOL_20161008.pdf

Stage of review

Review Ongoing

Subject index terms status

Subject indexing assigned by CRD

Subject index terms

Dementia; Demography; Epidemiology; Geography; Humans

Date of registration in PROSPERO

08 November 2016

Date of first submission

08 December 2020

Details of any existing review of the same topic by the same authors

This review is an update of the earlier systematic review and is being undertaken in the light of the publication of further

new research which will assist in looking at the role of geographical location has dementia risk.

The citation for the existing review:

Russ, T.C., Batty, G.B., Hearnshaw, G.F., Fenton, C., Starr, J.M. Geographical variation in dementia: systematic review with meta-analysis. International Journal of Epidemiology 2012;41:1012-1032 . <https://doi.org/10.1093/ije/dys103>

Stage of review at time of this submission [1 change]

Stage	Started	Completed
Preliminary searches	Yes	No
Piloting of the study selection process	Yes	No
Formal screening of search results against eligibility criteria	Yes	No
Data extraction	Yes	No
Risk of bias (quality) assessment	Yes	No
Data analysis	Yes	No

Revision note

7th December 2020: This revision is to amend the completion date to July 2021 and to add additional authors to the review due to the scale of the review. I have also updated the parts of the review that are completed/underway

The record owner confirms that the information they have supplied for this submission is accurate and complete and they understand that deliberate provision of inaccurate information or omission of data may be construed as scientific misconduct.

The record owner confirms that they will update the status of the review when it is completed and will add publication details in due course.

Versions

08 November 2016

08 December 2020

31 July 2021

Appendix 2

This Appendix relates to [Chapter 2: Systematic review and meta-analysis of rural or urban living with risk of dementia](#). The RECORD checklist for the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement [[433](#), [751](#)] is included below.

This is included with permission under Creative Commons (CC BY) Licence 4.0 <https://creativecommons.org/licenses/by/4.0/>. No changes to the original were made. Source: PLoS Med. 2015;12(10):e1001885. Authors: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM; RECORD Working Committee. Title: “The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement.” Available at <https://doi.org/10.1371/journal.pmed.1001885>, 2015.

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			
Objectives	3	State specific objectives, including any prespecified hypotheses			
Methods					
Study Design	4	Present key elements of study design early in the paper			
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			

Participants	6	(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			

Bias	9	Describe any efforts to address potential sources of bias			
Study size	10	Explain how the study size was arrived at			
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (<i>e.g.</i> , numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (<i>i.e.</i> , study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	
Descriptive data	14	(a) Give characteristics of study participants (<i>e.g.</i> , demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) <i>Cohort study</i> - summarise follow-up time (<i>e.g.</i> , average and total amount)			
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure			

		category, or summary measures of exposure <i>Cross-sectional study</i> - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			
Discussion					
Key results	18	Summarise key results with reference to study objectives			
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	

*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.

Appendix 3

This Appendix related to [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#). Random Rounding to the base of 3 (RR3) is used for outputs from Aotearoa New Zealand's Integrated Data Infrastructure (IDI). Statistics New Zealand produce an information sheet on RR3 [\[632\]](#). I enclose the sheet as a pdf in this Appendix for information.

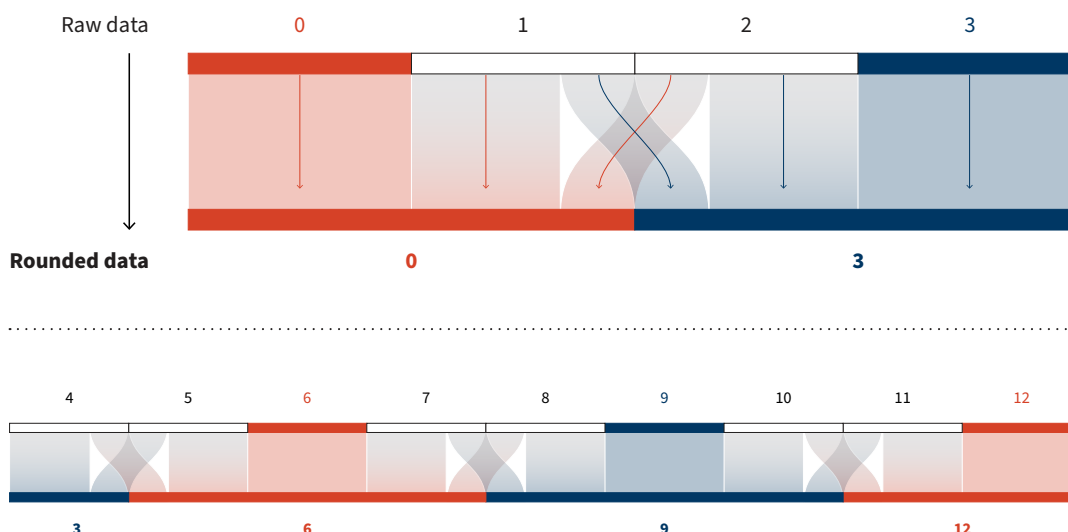
Random Rounding

What does this method do?

We can protect information in counts tables by random rounding to base 3 (RR3). The counts are randomly rounded to base three in a consistent manner. This is to disguise small counts, but all cells in the table are randomly rounded. The effect is to make the output more confidential, by generally preventing individuals' data from being released.

How does this affect the final data?

For small numbers, where there is the most risk that individuals could be identified, there are larger percentage changes compared with larger numbers. For example, a cell with a one changed to a three has been changed by 200 percent, but a cell with 1,001 changed to 1,002 has been changed by only 0.1 percent. When analysing data, small counts need to be treated with caution but for the larger values the percentage changes in these cells do not cause a problem.



Appendix 4

This Appendix related to [Chapter 4: Investigating dementia in Aotearoa New Zealand \(ANZ\) using routinely collected data](#). The datasets used for the boundaries in Figure [4.4 ANZ map with District Health Boards 2015](#) provided courtesy of Jessie Colbert, University of Auckland for the 2015 representation of Aotearoa New Zealand (ANZ) map with district health boards and the "*zoomed in*" view of Auckland Central Business District with different layers of geographical scales:

CAU

- Stats NZ area unit 2013 shapefile
- <https://datafinder.stats.govt.nz/layer/25743-area-unit-2013/>
- These boundaries align with the NZ 2013 Census.
- CAUs are no longer used for the NZ Census post the 2013 Census and have been replaced with Statistical Area 2 (SA2) units.
- CAUs were updated up to 2017, but beyond then have been discontinued.

Meshblock (MB)

- Stats NZ meshblock 2013 shapefile
- <https://datafinder.stats.govt.nz/layer/8347-meshblock-2013/>
- These boundaries align with the NZ 2013 Census (the spatial units can be linked to the 2013 Census results).
- MBs have been updated every year since 2013, however are no longer the primary spatial unit of choice for the NZ Census, they have been replaced longer term with Statistical Area 1 (SA1) units.
- MBs are however still used as the building blocks of the Stats NZ geographic hierarchy: <https://www.stats.govt.nz/methods/geographic-hierarchy/>

Data Zones (DZ) 2013

- <https://hgd.auckland.ac.nz/data/#IMD13-data>
- The DZs were first created for the NZIMD 2013.

- <https://hgd.auckland.ac.nz/datazones/>
- There is an updated version of the DZs for 2018 aligning with the release of the NZIMD 2018.

District Health Board (DHB)

- Stats NZ district health board 2015 shapefile
- <https://datafinder.stats.govt.nz/layer/87883-district-health-board-2015/>
- DHBs are based on the Stats NZ Meshblocks.
- The 20 DHBs were formally disestablished in 2022 and replaced by new Health NZ boundaries.

Appendix 5

Detailed processes of extraction first diagnosis of dementia

This Appendix relates to [Chapter 5: Methodology for the SMS1947 dementia data linkage study](#). I outline additional steps in the Methods required within the *SMS1947 dementia data linkage study* in [Chapter 6: Results of the SMS1947 dementia data linkage study](#) to ensure the first diagnosis of dementia was obtained from each of the datasets (SMR 01, SMR 04, PIS or NRS mortality). ‘Collapsing records’ refers to reducing the number of records an individual has in a large dataset to one record per individual with the key information relevant to the research question. An individual may feature many times in an individual dataset with multiple entries for hospital discharges or drug prescriptions. It is important to have a ‘core’ dataset which contains specific information for each individual. To do this the many records need ‘collapsed’ down to one record with the key information required.

SMR Datasets

Within the SMR datasets, I *flagged* dementia within the individual SMR 01 or SMR 04 datasets and then merged them into one *flagged* (dementia) SMR dataset. Table A.1 outlines the steps. Dementia was coded as 1 when *flagged* in the dataset.

Step	Process
1	All with dementia were flagged within the dataset using ICD 9 and 10 codes
2	I created a 'first_age' column and inputted the discharge date in months
3	I created a source column referring to whether it was SMR 01 or SMR 04 as the original source of dementia diagnosis when the SMR datasets were merged
4	I removed columns/variables not required
5	I collapsed down the data file from records to individuals (so one person had one entry) by a) grouping a person's individual ID, b) filtering to only those with dementia, and then c) ascending in order from youngest 'first_age' and selecting the first record entry (their first dementia mentioned with corresponding age) for each individual
6	The dataset was referred to 'dementia_smr' in the R code script

Table A.1: Process for extracting first diagnosis of dementia in SMR

NRS Mortality Dataset

For the mortality dataset, if dementia was present in an ICD code, then there was only one recording of this, i.e., date of death. This meant there was only one record per individual. Therefore, unlike the other two datasets above, I did not have to collapse many records from each individual down to one record.

Step	Process
1	All with dementia were flagged within the dataset using ICD 9 and 10 codes
2	I created a 'first_age' column and inputted the age of death in months
3	I created a source column and filled it with 'Deaths' so I knew the original source of the dementia diagnosis when merged with the other datasets
4	I removed columns/variables not required
5	I only had one record for each individual in this dataset given it was deaths data and only possible to have one entry. Therefore, this date (age of death in months) was used as 'first_age' for those with dementia recorded.
6	The dataset was referred to 'dementia_deaths' in the R code script

Table A.2: Process for extracting first diagnosis of dementia in NRS mortality

PIS Dataset

The PIS dataset has a number of records for each individual. A person will be prescribed anti-dementia drugs many times across their life-course on repeat prescriptions. An individual might also be on more than one anti-dementia drug across the time of linked data in the National Safe Haven, so obtaining the record of the *first* anti-dementia drug is important to represent the *earliest* entry in the datasets as a proxy for their *first* diagnosis of dementia. Everyone on anti-dementia drugs as listed before in [Anti-dementia drugs used for dementia ascertainment in Scotland](#) was flagged. The next step was to order these flagged individuals with a *first date* for the prescription. The process and steps are outlined in Table A.3.

Step	Process
1	All with dementia were flagged within the dataset using the generic names (and cross-checked with BNF codes) for anti-dementia drugs used specifically for dementia
2	I created a ‘first_age’ column and inputted the prescribed age in months
3	I created a source column and filled it with ‘PIS so I knew the original source of the dementia diagnosis when merged with the other datasets
4	I removed columns/variables not required
5	I collapsed down the data file from records to individuals (so one person had one entry) by a) grouping a person’s individual ID, b) filtering to only those with dementia, and then c) ascending in order from youngest ‘first_age’ and selecting the first record entry (their first anti-dementia drug prescribed with corresponding age) for each individual
6	The dataset was referred to ‘dementia_drugs’ in the R code script

Table A.3: Process for extracting first dementia diagnosis in PIS

Appendix 6

This Appendix relates to [Chapter 6: Results of the SMS1947 dementia data linkage study](#). These are additional results. A Cox Proportional Hazard model investigating ‘Accessible’ versus ‘Remote’ areas was undertaken. Accessible areas were grouped together and compared against all Remote areas grouped together. Accessible areas included both urban and rural areas: Urban classifications 1,2 and 3 (1=large urban areas; 2=other urban areas; 3=accessible small towns; 6=accessible rural areas). Remote areas included both urban and rural areas: Urban classifications 4 and 5, along with Rural classifications 7 and 8 (4=remote small towns; 5=very remote small towns; 7=remote rural areas; 8=very remote rural areas).

	Urban-Rural classification (levels)	Size (<i>n</i>)	Hazard Ratio	Lower 95% CI	Upper 95% CI
Model 1	Accessible (1,2,3 and 6)		1 (Ref.)		
(unadjusted)	Remote (4,5,7 and 8)		0.90	0.79	1.01
Model 2	Accessible (1,2,3 and 6)		1 (Ref.)		
(sex)	Remote (4,5,7 and 8)		0.90	0.79	1.01
Model 3	Accessible (1,2,3 and 6)		1 (Ref.)		
(sex, intelligence)	Remote (4,5,7 and 8)		0.92	0.81	1.04
Model 4	Accessible (1,2,3 and 6)		1 (Ref.)		
(sex, SIMD)	Remote (4,5,7 and 8)		0.91	0.80	1.03
Model 5	Accessible (1,2,3 and 6)		1 (Ref.)		
(sex, intelligence, SIMD)	Remote (4,5,7 and 8)		0.92	0.81	1.04

Table A.4: Cox Proportional Hazard Models comparing all accessible areas (urban and rural) grouped together, versus all remote areas (urban and rural) classifications.

Accessible is the reference. Accessible areas grouped together is the reference (Accessible=1, no difference). Accessible areas include Urban classifications 1,2 and 3 (1=large urban areas; 2=other urban areas; 3=accessible small towns; 6=accessible rural areas), and Remote areas include Urban classifications 4 and 5, along with Rural classifications 7 and 8 (4=remote small towns; 5=very remote small towns; 7=remote rural areas; 8=very remote rural areas). Intelligence refers to the corrected and standardised Moray House Test, MHT, (standardised to a mean =0, SD = 1, for all SMS1947 participants), SIMD refers to the Scottish Index of Multiple Deprivation variable in deciles. Model 1 unadjusted; Model 2 adjusted for sex; Model 3 adjusted for sex and intelligence; Model 4 adjusted for sex and SIMD; Model 5 adjusted for sex, intelligence and SIMD.

Bibliography

- [1] K.E. Walesby et al. “Geographical variation in dementia: systematic review with meta-analysis”. In: *medRxiv* (Dec. 2023). URL: <https://doi.org/10.1101/2023.12.11.23299178>.
- [2] KE Walesby, JK Harrison, and TC Russ. “What big data could achieve in Scotland”. In: *Journal of the Royal College of Physicians of Edinburgh* 47.2 (Mar. 2017), pp. 1–6. DOI: [10.4997/JRCPE.2017.201](https://doi.org/10.4997/JRCPE.2017.201).
- [3] Katherine Elizabeth Walesby et al. “Prevalence and geographical variation of dementia in New Zealand from 2012 to 2015: Brief report utilising routinely collected data within the Integrated Data Infrastructure”. In: *Australasian Journal on Ageing* 39.3 (Sept. 2020), pp. 297–304. DOI: [10.1111/ajag.12790](https://doi.org/10.1111/ajag.12790).
- [4] Daniel J. Exeter and Katherine E. Walesby. “Using big data in primary care research”. In: *How to Do Primary Care Research*. CRC Press, Oct. 2018, pp. 65–73. DOI: [10.1201/9781351014519-8](https://doi.org/10.1201/9781351014519-8).
- [5] Māori proverbial saying. *Whāia te iti kahurangi! Strive for something of great value!* URL: <https://kupu.maori.nz/kupu/Wh%C4%81ia-te-iti-kahurangi!#:~:text=Strive%20for%20something%20of%20great%20value!%20This%20comes>.
- [6] Jane A H Masoli et al. “New horizons in the role of digital data in the healthcare of older people”. In: *Age and Ageing* 52.8 (Aug. 2023). DOI: [10.1093/ageing/afad134](https://doi.org/10.1093/ageing/afad134).
- [7] Oliver M Todd et al. “New Horizons in the use of routine data for ageing research”. In: *Age and Ageing* 49.5 (Aug. 2020), pp. 716–722. DOI: [10.1093/ageing/afaa018](https://doi.org/10.1093/ageing/afaa018).
- [8] ASF Doney et al. “Investigating the Relationship Between Type 2 Diabetes and Dementia Using Electronic Medical Records in the GoDARTS Bioresource”. In: *Diabetes Care* 42.10 (Oct. 2019), pp. 1973–1980. DOI: [10.2337/dc19-0380](https://doi.org/10.2337/dc19-0380).
- [9] JM. Starr and KE. Walesby. “Diagnosis and Management of Dementia in older people”. In: *Medicine* 45.1 (2017), pp. 51–55.
- [10] John M. Starr and Katherine E. Walesby. “Vascular and Mixed Dementia”. In: *Encyclopedia of Geropsychology*. Singapore: Springer Singapore, 2016, pp. 1–8. DOI: [10.1007/978-981-287-080-3_313-1](https://doi.org/10.1007/978-981-287-080-3_313-1).
- [11] Carly Welch et al. “Development of a UK core dataset for geriatric medicine research: a position statement and results from a Delphi consensus process”. In: *BMC Geriatrics* 23.1 (Mar. 2023), p. 168. DOI: [10.1186/s12877-023-03805-5](https://doi.org/10.1186/s12877-023-03805-5).
- [12] Stephen Lim et al. “Engagement in research during specialist geriatric medicine training: results of a national trainee survey.” In: *Clinical medicine (London, England)* 22.6 (Nov. 2022), pp. 553–558. DOI: [10.7861/clinmed.2022-0283](https://doi.org/10.7861/clinmed.2022-0283).
- [13] Lorna Caulfield et al. “Effect of Angiotensin System Inhibitors on Physical Performance in Older People – A Systematic Review and Meta-Analysis”. In: *Journal of the American Medical Directors Association* 22.6 (June 2021), pp. 1215–1221. DOI: [10.1016/j.jamda.2020.07.012](https://doi.org/10.1016/j.jamda.2020.07.012).
- [14] Carly Welch et al. “Growing research in geriatric medicine: a trainee perspective”. In: *Age and Ageing* 49.5 (Aug. 2020), pp. 733–737. DOI: [10.1093/ageing/afaa052](https://doi.org/10.1093/ageing/afaa052).

- [15] JK Burton et al. “Predicting Discharge to Institutional Long-Term Care After Stroke: A Systematic Review and Metaanalysis”. In: *Journal of the American Geriatrics Society* 66.1 (Jan. 2018), pp. 161–169. DOI: [10.1111/jgs.15101](https://doi.org/10.1111/jgs.15101).
- [16] Jennifer Kirsty Harrison et al. “Predicting discharge to institutional long-term care following acute hospitalisation: a systematic review and meta-analysis”. In: *Age and Ageing* 46.4 (July 2017), pp. 547–558. DOI: [10.1093/ageing/afx047](https://doi.org/10.1093/ageing/afx047).
- [17] Arabella L Simpkin and Katherine E Walesby. “Training tomorrow’s doctors”. In: *Future Hospital Journal* 4.1 (Feb. 2017), pp. 56–60. DOI: [10.7861/futurehosp.4-1-56](https://doi.org/10.7861/futurehosp.4-1-56).
- [18] KE Walesby et al. “Valuing Our Trainees: The Future of Medicine in the UK”. In: *Journal of the Royal College of Physicians of Edinburgh* 46.3 (Sept. 2016), pp. 146–146. DOI: [10.4997/jrcpe.2016.301](https://doi.org/10.4997/jrcpe.2016.301).
- [19] F Assal. “History of Dementia”. In: *Frontiers in Neurology and Neuroscience* 44 (2019), pp. 118–126. DOI: [10.1159/000494959](https://doi.org/10.1159/000494959).
- [20] World Health Organisation. *Dementia: Key facts*. 2024. URL: <https://www.who.int/news-room/fact-sheets/detail/dementia>.
- [21] World Health Organisation. *World Health Organisation: International Classification of Diseases (ICD) 11th Edition*. Geneva: WHO, 2019. URL: <https://icd.who.int/en>.
- [22] World Health Organisation. *World Health Organisation: Dementia - key factors*. URL: <https://www.who.int/news-room/fact-sheets/detail/dementia>.
- [23] Z Arvanitakis, RC Shah, and DA Bennett. “Diagnosis and Management of Dementia: Review”. In: *JAMA* 322.16 (Oct. 2019), p. 1589. DOI: [10.1001/jama.2019.4782](https://doi.org/10.1001/jama.2019.4782).
- [24] World Health Organisation. *Dementia*. 2024. URL: https://www.who.int/health-topics/dementia#tab=tab_1.
- [25] Guy McKhann et al. “Clinical diagnosis of Alzheimer’s disease”. In: *Neurology* 34.7 (July 1984), pp. 939–939. DOI: [10.1212/WNL.34.7.939](https://doi.org/10.1212/WNL.34.7.939).
- [26] Guy M. McKhann et al. “The diagnosis of dementia due to Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease”. In: *Alzheimer’s & Dementia* 7.3 (May 2011), pp. 263–269. DOI: [10.1016/j.jalz.2011.03.005](https://doi.org/10.1016/j.jalz.2011.03.005).
- [27] JCS Breitner. “Dementia—Epidemiological Considerations, Nomenclature, and a Tacit Consensus Definition”. In: *Journal of Geriatric Psychiatry and Neurology* 19.3 (Sept. 2006), pp. 129–136. DOI: [10.1177/0891988706291081](https://doi.org/10.1177/0891988706291081).
- [28] Chiara C. Brück et al. “Heterogeneity in Reports of Dementia Disease Duration and Severity: A Review of the Literature”. In: *Journal of Alzheimer’s Disease* 84.4 (Dec. 2021), pp. 1515–1522. DOI: [10.3233/JAD-210544](https://doi.org/10.3233/JAD-210544).
- [29] Clive G. Ballard. “Definition and Diagnosis of Dementia with Lewy Bodies”. In: *Dementia and Geriatric Cognitive Disorders* 17.Suppl. 1 (2004), pp. 15–24. DOI: [10.1159/000074678](https://doi.org/10.1159/000074678).
- [30] H. Loo and J. Plas. “Dementia – A Semantic Definition”. In: *Gerontology* 32.1 (1986), pp. 64–66. DOI: [10.1159/000212831](https://doi.org/10.1159/000212831).

- [31] A Wallin. “Current definition and classification of dementia diseases”. In: *Acta Neurologica Scandinavica* 94 (Dec. 1996), pp. 39–44. DOI: [10.1111/j.1600-0404.1996.tb00371.x](https://doi.org/10.1111/j.1600-0404.1996.tb00371.x).
- [32] Bradley F Boeve et al. “Advances and controversies in frontotemporal dementia: diagnosis, biomarkers, and therapeutic considerations”. In: *The Lancet Neurology* 21.3 (Mar. 2022), pp. 258–272. DOI: [10.1016/S1474-4422\(21\)00341-0](https://doi.org/10.1016/S1474-4422(21)00341-0).
- [33] GaryW. Small and LissyF. Jarvik. “THE DEMENTIA SYNDROME”. In: *The Lancet* 320.8313 (Dec. 1982), pp. 1443–1446. DOI: [10.1016/S0140-6736\(82\)91339-3](https://doi.org/10.1016/S0140-6736(82)91339-3).
- [34] Ravi Bhat and Kenneth Rockwood. “Psychiatric Complications of Dementia”. In: *The Canadian Journal of Psychiatry* 56.7 (July 2011), pp. 398–407. DOI: [10.1177/070674371105600703](https://doi.org/10.1177/070674371105600703).
- [35] R Zimmer and H Lauter. “Diagnosis, Differential Diagnosis and Nosologic Classification of the Dementia Syndrome”. In: *Pharmacopsychiatry* 21.S 1 (Mar. 1988), pp. 1–7. DOI: [10.1055/s-2007-1017057](https://doi.org/10.1055/s-2007-1017057).
- [36] A Robles Bayón. “Degenerative dementias: a question of syndrome or disease?” In: *Neurología (English Edition)* 37.6 (July 2022), pp. 480–491. DOI: [10.1016/j.nrleng.2019.03.027](https://doi.org/10.1016/j.nrleng.2019.03.027).
- [37] Frank Jessen and Lutz Frölich. “ICD-11: Neurokognitive Störungen”. In: *Fortschritte der Neurologie · Psychiatrie* 86.03 (Mar. 2018), pp. 172–177. DOI: [10.1055/s-0044-101607](https://doi.org/10.1055/s-0044-101607).
- [38] Christian Behl. *Alzheimer’s Disease Research What Has Guided Research So Far and Why It Is High Time for a Paradigm Shift*. Tech. rep.
- [39] Olivia Sheppard and Michael Coleman. “Alzheimer’s Disease: Etiology, Neuropathology and Pathogenesis”. In: *Alzheimer’s Disease: Drug Discovery*. Exon Publications, Dec. 2020, pp. 1–22. DOI: [10.36255/exonpublications.alzheimersdisease.2020.ch1](https://doi.org/10.36255/exonpublications.alzheimersdisease.2020.ch1).
- [40] Hamzah Mohammed. “Recognizing African-American contributions to neurology: The role of Solomon Carter Fuller (1872–1953) in Alzheimer’s disease research”. In: *Alzheimer’s & Dementia* 17.2 (Feb. 2021), pp. 246–250. DOI: [10.1002/alz.12183](https://doi.org/10.1002/alz.12183).
- [41] M Sarazin et al. “The path to next-generation disease-modifying immunomodulatory combination therapies in Alzheimer’s disease”. In: *Nature Aging* 4.6 (June 2024), pp. 761–770. DOI: [10.1038/s43587-024-00630-2](https://doi.org/10.1038/s43587-024-00630-2).
- [42] Eric M Reiman et al. “A chance to prevent Alzheimer’s disease sooner than you think”. In: *The Lancet Neurology* 23.2 (Feb. 2024), pp. 144–145. DOI: [10.1016/S1474-4422\(23\)00508-2](https://doi.org/10.1016/S1474-4422(23)00508-2).
- [43] Joseph L. Price et al. “Neuropathology of nondemented aging: Presumptive evidence for preclinical Alzheimer disease”. In: *Neurobiology of Aging* 30.7 (July 2009), pp. 1026–1036. DOI: [10.1016/j.neurobiolaging.2009.04.002](https://doi.org/10.1016/j.neurobiolaging.2009.04.002).
- [44] Dora Games et al. “Alzheimer-type neuropathology in transgenic mice overexpressing V717F β -amyloid precursor protein”. In: *Nature* 373.6514 (Feb. 1995), pp. 523–527. DOI: [10.1038/373523a0](https://doi.org/10.1038/373523a0).

- [45] Kasper P Kepp et al. “The amyloid cascade hypothesis: an updated critical review”. In: *Brain* 146.10 (Oct. 2023), pp. 3969–3990. DOI: [10.1093/brain/awad159](https://doi.org/10.1093/brain/awad159).
- [46] Aamira J. Huq et al. “Genetic testing in dementia-A medical genetics perspective”. In: *International Journal of Geriatric Psychiatry* 36.8 (Aug. 2021), pp. 1158–1170. DOI: [10.1002/gps.5535](https://doi.org/10.1002/gps.5535).
- [47] Percy Griffin et al. “Developments in understanding early onset Alzheimer’s disease”. In: *Alzheimer’s & Dementia* 19.S9 (Nov. 2023). DOI: [10.1002/alz.13353](https://doi.org/10.1002/alz.13353).
- [48] Clement T Loy et al. “Genetics of dementia”. In: *The Lancet* 383.9919 (Mar. 2014), pp. 828–840. DOI: [10.1016/S0140-6736\(13\)60630-3](https://doi.org/10.1016/S0140-6736(13)60630-3).
- [49] Amalia C. Bruni, Maria E. Conidi, and Livia Bernardi. “Genetics in Degenerative Dementia”. In: *Alzheimer Disease & Associated Disorders* 28.3 (July 2014), pp. 199–205. DOI: [10.1097/WAD.0000000000000046](https://doi.org/10.1097/WAD.0000000000000046).
- [50] Liyong Wu et al. “Early-Onset Familial Alzheimer’s Disease (EOFAD)”. In: *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques* 39.4 (July 2012), pp. 436–445. DOI: [10.1017/S0317167100013949](https://doi.org/10.1017/S0317167100013949).
- [51] Samantha M Loi, Monica Cations, and Dennis Velakoulis. “Young-onset dementia diagnosis, management and care: a narrative review”. In: *Medical Journal of Australia* 218.4 (Mar. 2023), pp. 182–189. DOI: [10.5694/mja2.51849](https://doi.org/10.5694/mja2.51849).
- [52] J Farlow and T Foroud. “The Genetics of Dementia”. In: *Seminars in Neurology* 33.04 (Nov. 2013), pp. 417–422. DOI: [10.1055/s-0033-1359313](https://doi.org/10.1055/s-0033-1359313).
- [53] Martin N Rossor et al. “The diagnosis of young-onset dementia”. In: *The Lancet Neurology* 9.8 (Aug. 2010), pp. 793–806. DOI: [10.1016/S1474-4422\(10\)70159-9](https://doi.org/10.1016/S1474-4422(10)70159-9).
- [54] Alberto Serrano-Pozo, Sudeshna Das, and Bradley T Hyman. “APOE and Alzheimer’s disease: advances in genetics, pathophysiology, and therapeutic approaches”. In: *The Lancet Neurology* 20.1 (Jan. 2021), pp. 68–80. DOI: [10.1016/S1474-4422\(20\)30412-9](https://doi.org/10.1016/S1474-4422(20)30412-9).
- [55] Ana-Caroline Raulin et al. “ApoE in Alzheimer’s disease: pathophysiology and therapeutic strategies”. In: *Molecular Neurodegeneration* 17.1 (Nov. 2022), p. 72. DOI: [10.1186/s13024-022-00574-4](https://doi.org/10.1186/s13024-022-00574-4).
- [56] Yuka A. Martens et al. “ApoE Cascade Hypothesis in the pathogenesis of Alzheimer’s disease and related dementias”. In: *Neuron* 110.8 (Apr. 2022), pp. 1304–1317. DOI: [10.1016/j.neuron.2022.03.004](https://doi.org/10.1016/j.neuron.2022.03.004).
- [57] Chia-Chen Liu et al. “Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy”. In: *Nature Reviews Neurology* 9.2 (Feb. 2013), pp. 106–118. DOI: [10.1038/nrneurol.2012.263](https://doi.org/10.1038/nrneurol.2012.263).
- [58] E. H. Corder et al. “Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer’s Disease in Late Onset Families”. In: *Science* 261.5123 (Aug. 1993), pp. 921–923. DOI: [10.1126/science.8346443](https://doi.org/10.1126/science.8346443).
- [59] Jean-Charles Lambert et al. “Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease”. In: *Nature Genetics* 45.12 (Dec. 2013), pp. 1452–1458. DOI: [10.1038/ng.2802](https://doi.org/10.1038/ng.2802).

- [60] BW Kunkle et al. “Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates A β , tau, immunity and lipid processing”. In: *Nature Genetics* 51.3 (Mar. 2019), pp. 414–430. DOI: [10.1038/s41588-019-0358-2](https://doi.org/10.1038/s41588-019-0358-2).
- [61] Ditte Demontis et al. “Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains”. In: *Nature Genetics* 55.2 (Feb. 2023), pp. 198–208. DOI: [10.1038/s41588-022-01285-8](https://doi.org/10.1038/s41588-022-01285-8).
- [62] Jesús Andrade-Guerrero et al. “Alzheimer’s Disease: An Updated Overview of Its Genetics”. In: *International Journal of Molecular Sciences* 24.4 (Feb. 2023), p. 3754. DOI: [10.3390/ijms24043754](https://doi.org/10.3390/ijms24043754).
- [63] Lucia Migliore and Fabio Coppedè. “Gene–environment interactions in Alzheimer disease: the emerging role of epigenetics”. In: *Nature Reviews Neurology* 18.11 (Nov. 2022), pp. 643–660. DOI: [10.1038/s41582-022-00714-w](https://doi.org/10.1038/s41582-022-00714-w).
- [64] M. Ryan Corces et al. “Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer’s and Parkinson’s diseases”. In: *Nature Genetics* 52.11 (Nov. 2020), pp. 1158–1168. DOI: [10.1038/s41588-020-00721-x](https://doi.org/10.1038/s41588-020-00721-x).
- [65] Bartosz Słowikowski et al. “Epigenetics and the neurodegenerative process”. In: *Epigenomics* 16.7 (Apr. 2024), pp. 473–491. DOI: [10.2217/epi-2023-0416](https://doi.org/10.2217/epi-2023-0416).
- [66] Xushen Xiong et al. “Epigenomic dissection of Alzheimer’s disease pinpoints causal variants and reveals epigenome erosion”. In: *Cell* 186.20 (Sept. 2023), pp. 4422–4437. DOI: [10.1016/j.cell.2023.08.040](https://doi.org/10.1016/j.cell.2023.08.040).
- [67] Abhirami Ramakrishnan et al. “Epigenetic dysregulation in Alzheimer’s disease peripheral immunity”. In: *Neuron* 112.8 (Apr. 2024), pp. 1235–1248. DOI: [10.1016/j.neuron.2024.01.013](https://doi.org/10.1016/j.neuron.2024.01.013).
- [68] Chiara Fenoglio et al. “Role of Genetics and Epigenetics in the Pathogenesis of Alzheimer’s Disease and Frontotemporal Dementia”. In: *Journal of Alzheimer’s Disease* 62.3 (Mar. 2018), pp. 913–932. DOI: [10.3233/JAD-170702](https://doi.org/10.3233/JAD-170702).
- [69] Sonia Mayo et al. “Recent Evidence in Epigenomics and Proteomics Biomarkers for Early and Minimally Invasive Diagnosis of Alzheimer’s and Parkinson’s Diseases”. In: *Current Neuropharmacology* 19.8 (Aug. 2021), pp. 1273–1303. DOI: [10.2174/1570159X19666201223154009](https://doi.org/10.2174/1570159X19666201223154009).
- [70] Paul Peixoto et al. “From 1957 to Nowadays: A Brief History of Epigenetics”. In: *International Journal of Molecular Sciences* 21.20 (Oct. 2020), p. 7571. DOI: [10.3390/ijms21207571](https://doi.org/10.3390/ijms21207571).
- [71] C Grunau et al. “The Epigenetics Dilemma”. In: *Genes* 11.1 (Dec. 2019), p. 23. DOI: [10.3390/genes11010023](https://doi.org/10.3390/genes11010023).
- [72] Caroline Ospelt. “A brief history of epigenetics”. In: *Immunology Letters* 249 (Sept. 2022), pp. 1–4. DOI: [10.1016/j.imlet.2022.08.001](https://doi.org/10.1016/j.imlet.2022.08.001).
- [73] Andrea Stoccoro and Fabio Coppedè. “Role of Epigenetics in Alzheimer’s Disease Pathogenesis”. In: *Neurodegenerative Disease Management* 8.3 (June 2018), pp. 181–193. DOI: [10.2217/nmt-2018-0004](https://doi.org/10.2217/nmt-2018-0004).
- [74] Frank J. Wolters and M. Arfan Ikram. “Epidemiology of Vascular Dementia”. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 39.8 (Aug. 2019), pp. 1542–1549. DOI: [10.1161/ATVBAHA.119.311908](https://doi.org/10.1161/ATVBAHA.119.311908).

- [75] Shyamal C. Bir et al. “Emerging Concepts in Vascular Dementia: A Review”. In: *Journal of Stroke and Cerebrovascular Diseases* 30.8 (Aug. 2021), p. 105864. DOI: [10.1016/j.jstrokecerebrovasdis.2021.105864](https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105864).
- [76] B Grandal Leiros et al. “Prevalencia y concordancia entre diagnóstico clínico y anatomopatológico de demencia en una clínica psicogeriatrica”. In: *Neurologia* 33.1 (2018), pp. 13–17. DOI: [10.1016/j.nrl.2016.04.011](https://doi.org/10.1016/j.nrl.2016.04.011).
- [77] Alzheimer Europe. *Estimating the prevalence of dementia in Europe*. Tech. rep. Alzheimer Europe, 2020. URL: https://www.alzheimer-europe.org/sites/default/files/alzheimer_europe_dementia_in_europe_yearbook_2019.pdf.
- [78] Julie A. Schneider et al. “Mixed brain pathologies account for most dementia cases in community-dwelling older persons”. In: *Neurology* 69.24 (Dec. 2007), pp. 2197–2204. DOI: [10.1212/01.wnl.0000271090.28148.24](https://doi.org/10.1212/01.wnl.0000271090.28148.24).
- [79] Emma Nichols et al. “Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019”. In: *The Lancet Public Health* 7.2 (Feb. 2022), e105–e125. DOI: [10.1016/S2468-2667\(21\)00249-8](https://doi.org/10.1016/S2468-2667(21)00249-8).
- [80] M Prince et al. “The global prevalence of dementia: A systematic review and metaanalysis”. In: *Alzheimer’s & Dementia* 9.1 (Jan. 2013), p. 63. DOI: [10.1016/j.jalz.2012.11.007](https://doi.org/10.1016/j.jalz.2012.11.007).
- [81] CP Ferri et al. “Global prevalence of dementia: a Delphi consensus study”. In: *The Lancet* 366.9503 (Dec. 2005), pp. 2112–2117. DOI: [10.1016/S0140-6736\(05\)67889-0](https://doi.org/10.1016/S0140-6736(05)67889-0).
- [82] Authors Martin Prince et al. *World Alzheimer Report 2015 The Global Impact of Dementia An Analysis of prevalence, Incidence, cost And Trends*. Tech. rep. URL: www.alz.co.uk/worldreport2015corrections.
- [83] Global Council on Brain Health. *Global Council on Brain Health (GCBH) Summary*. Tech. rep.
- [84] Yu Tzu Wu et al. *Dementia in western Europe: Epidemiological evidence and implications for policy making*. Jan. 2016. DOI: [10.1016/S1474-4422\(15\)00092-7](https://doi.org/10.1016/S1474-4422(15)00092-7).
- [85] Martin Prince et al. “Recent global trends in the prevalence and incidence of dementia, and survival with dementia”. In: *Alzheimer’s Research & Therapy* 8.1 (Dec. 2016), p. 23. DOI: [10.1186/s13195-016-0188-8](https://doi.org/10.1186/s13195-016-0188-8).
- [86] Ingmar Skoog. “Dementia incidence — the times, they are a-changing”. In: *Nature Reviews Neurology* 12.6 (June 2016), pp. 316–318. DOI: [10.1038/nrneurol.2016.55](https://doi.org/10.1038/nrneurol.2016.55).
- [87] Yu-Tzu Wu et al. “The changing prevalence and incidence of dementia over time — current evidence”. In: *Nature Reviews Neurology* 13.6 (June 2017), pp. 327–339. DOI: [10.1038/nrneurol.2017.63](https://doi.org/10.1038/nrneurol.2017.63).
- [88] Mary Ganguli. “The times they are a-changin’: cohort effects in aging, cognition, and dementia”. In: *International Psychogeriatrics* 29.3 (Mar. 2017), pp. 353–355. DOI: [10.1017/S1041610216001927](https://doi.org/10.1017/S1041610216001927).
- [89] Kenneth M Langa. “Is the risk of Alzheimer’s disease and dementia declining?”. In: *Alzheimer’s Research & Therapy* 7.1 (Dec. 2015), p. 34. DOI: [10.1186/s13195-015-0118-1](https://doi.org/10.1186/s13195-015-0118-1).

- [90] Susanne Roehr et al. “Is dementia incidence declining in high-income countries? A systematic review and meta-analysis”. In: *Clinical Epidemiology* Volume 10 (Sept. 2018), pp. 1233–1247. DOI: [10.2147/CLEP.S163649](https://doi.org/10.2147/CLEP.S163649).
- [91] Claudia L. Satizabal et al. “Incidence of Dementia over Three Decades in the Framingham Heart Study”. In: *New England Journal of Medicine* 374.6 (Feb. 2016), pp. 523–532. DOI: [10.1056/NEJMoa1504327](https://doi.org/10.1056/NEJMoa1504327).
- [92] Alzheimer’s Disease International. *Dementia Statistics*. URL: <https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/>.
- [93] Dementia UK. *Dementia UK: Prevention and risk factors*. URL: <https://www.dementiuk.org/information-and-support/about-dementia/prevention-and-risk-factors/>.
- [94] Valery L Feigin et al. “Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019”. In: *The Lancet Neurology* 20.10 (Oct. 2021), pp. 795–820. DOI: [10.1016/S1474-4422\(21\)00252-0](https://doi.org/10.1016/S1474-4422(21)00252-0).
- [95] NHS. *Overview: Alzheimer’s disease*. URL: <https://www.nhs.uk/conditions/alzheimers-disease/>.
- [96] Scottish Government. *Dementia in Scotland: Everyone’s Story*. Tech. rep. 2023.
- [97] G Livingston et al. “Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission”. In: *The Lancet* 404.10452 (Aug. 2024), pp. 572–628. DOI: [10.1016/S0140-6736\(24\)01296-0](https://doi.org/10.1016/S0140-6736(24)01296-0).
- [98] World Health Organisation. *Global action plan on the public health response to dementia*. ISBN: 9789241513487. URL: <http://apps.who.int/bookorders..>
- [99] Jim Jackson et al. *Alzheimer Scotland: The Dementia Epidemic*. Tech. rep. Edinburgh: Alzheimer Scotland, June 2007, pp. 1–54. URL: https://www.alzscot.org/sites/default/files/images/0000/1945/Dementia_epidemic_web.pdf.
- [100] Horst Bickel Munich et al. *EUROCODE Workpackage 7 Prevalence*. Tech. rep.
- [101] Alzheimer Scotland. *Alzheimer Scotland: Statistics*. 2024. URL: <https://www.alzscot.org/our-work/campaigning-for-change/scotlands-national-dementia-strategy/statistics>.
- [102] Tom Snell and Robert Stewart. *Dementia UK THE FULL REPORT*. Tech. rep.
- [103] A HOFMAN et al. “The Prevalence of Dementia in Europe: A Collaborative Study of 1980–1990 Findings”. In: *International Journal of Epidemiology* 20.3 (1991), pp. 736–748. DOI: [10.1093/ije/20.3.736](https://doi.org/10.1093/ije/20.3.736).
- [104] GBD 2021 Demographics Collaborators. “Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021.” In: *Lancet (London, England)* 403.10440 (May 2024), pp. 1989–2056. DOI: [10.1016/S0140-6736\(24\)00476-8](https://doi.org/10.1016/S0140-6736(24)00476-8).
- [105] Michaël Schwarzingier and Carole Dufouil. “Forecasting the prevalence of dementia”. In: *The Lancet Public Health* 7.2 (Feb. 2022), e94–e95. DOI: [10.1016/S2468-2667\(21\)00277-2](https://doi.org/10.1016/S2468-2667(21)00277-2).

- [106] S. Long, C. Benoist, and W. Weidner. *World Alzheimer Report 2023 Reducing dementia risk: never too early, never too late*. Tech. rep. London: Alzheimer's Disease International, Sept. 2023, pp. 1–96.
- [107] Blazej Misiak et al. “European studies on the prevalence of dementia in the elderly: time for a step towards a methodological consensus”. In: *International Journal of Geriatric Psychiatry* 28.12 (Dec. 2013), pp. 1211–1221. DOI: [10.1002/gps.3948](https://doi.org/10.1002/gps.3948).
- [108] Katherine Leach and Allan Hicks. “The nurse’s role in closing the diagnostic gap for people with dementia”. In: *British Journal of Community Nursing* 18.9 (Sept. 2013), pp. 433–440. DOI: [10.12968/bjcn.2013.18.9.433](https://doi.org/10.12968/bjcn.2013.18.9.433).
- [109] Sube Banerjee. “A narrative review of evidence for the provision of memory services”. In: *International Psychogeriatrics* 27.10 (Oct. 2015), pp. 1583–1592. DOI: [10.1017/S1041610215000149](https://doi.org/10.1017/S1041610215000149).
- [110] Allison Lindauer et al. “Dementia 360 ECHO: Using technology to facilitate diagnosis and treatment”. In: *Gerontology & Geriatrics Education* 43.2 (Apr. 2022), pp. 202–208. DOI: [10.1080/02701960.2020.1835658](https://doi.org/10.1080/02701960.2020.1835658).
- [111] Daniel P. Perl. “Neuropathology of Alzheimer’s Disease”. In: *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* 77.1 (Jan. 2010), pp. 32–42. DOI: [10.1002/msj.20157](https://doi.org/10.1002/msj.20157).
- [112] Michael A. DeTure and Dennis W. Dickson. “The neuropathological diagnosis of Alzheimer’s disease”. In: *Molecular Neurodegeneration* 14.1 (Dec. 2019), p. 32. DOI: [10.1186/s13024-019-0333-5](https://doi.org/10.1186/s13024-019-0333-5).
- [113] Fanny M. Elahi and Bruce L. Miller. “A clinicopathological approach to the diagnosis of dementia”. In: *Nature Reviews Neurology* 13.8 (Aug. 2017), pp. 457–476. DOI: [10.1038/nrneurol.2017.96](https://doi.org/10.1038/nrneurol.2017.96).
- [114] Andrew E. Budson and Paul R. Solomon. “New Criteria for Alzheimer Disease and Mild Cognitive Impairment”. In: *The Neurologist* 18.6 (Nov. 2012), pp. 356–363. DOI: [10.1097/NRL.0b013e31826a998d](https://doi.org/10.1097/NRL.0b013e31826a998d).
- [115] Chloe Fawns-Ritchie and Ian J. Deary. “Reliability and validity of the UK Biobank cognitive tests”. In: *PLOS ONE* 15.4 (Apr. 2020), e0231627. DOI: [10.1371/journal.pone.0231627](https://doi.org/10.1371/journal.pone.0231627).
- [116] Marshal F. Folstein, Susan E. Folstein, and Paul R. McHugh. ““Mini-mental state””. In: *Journal of Psychiatric Research* 12.3 (Nov. 1975), pp. 189–198. DOI: [10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6).
- [117] H Brodaty et al. “The GPCOG: A New Screening Test for Dementia Designed for General Practice”. In: *Journal of the American Geriatrics Society* 50.3 (Mar. 2002), pp. 530–534. DOI: [10.1046/j.1532-5415.2002.50122.x](https://doi.org/10.1046/j.1532-5415.2002.50122.x).
- [118] Eneida Mioshi et al. “The Addenbrooke’s Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening”. In: *International Journal of Geriatric Psychiatry* 21.11 (Nov. 2006), pp. 1078–1085. DOI: [10.1002/gps.1610](https://doi.org/10.1002/gps.1610).
- [119] Vyara Valkanova and Klaus P. Ebmeier. “Neuroimaging in dementia”. In: *Maturitas* 79.2 (Oct. 2014), pp. 202–208. DOI: [10.1016/j.maturitas.2014.02.016](https://doi.org/10.1016/j.maturitas.2014.02.016).

- [120] Leonie Lampe et al. “Multiclass prediction of different dementia syndromes based on multi-centric volumetric MRI imaging”. In: *NeuroImage: Clinical* 37 (2023), p. 103320. DOI: [10.1016/j.nicl.2023.103320](https://doi.org/10.1016/j.nicl.2023.103320).
- [121] S Mahalingam and M-K Chen. “Neuroimaging in Dementias”. In: *Seminars in Neurology* 39.02 (Apr. 2019), pp. 188–199. DOI: [10.1055/s-0039-1678580](https://doi.org/10.1055/s-0039-1678580).
- [122] Maria Carmela Tartaglia, Howard J. Rosen, and Bruce L. Miller. “Neuroimaging in Dementia”. In: *Neurotherapeutics* 8.1 (Jan. 2011), pp. 82–92. DOI: [10.1007/s13311-010-0012-2](https://doi.org/10.1007/s13311-010-0012-2).
- [123] Harald Hampel et al. “Blood-based biomarkers for Alzheimer’s disease: Current state and future use in a transformed global healthcare landscape”. In: *Neuron* 111.18 (Sept. 2023), pp. 2781–2799. DOI: [10.1016/j.neuron.2023.05.017](https://doi.org/10.1016/j.neuron.2023.05.017).
- [124] Qingwei Ruan et al. “Emerging biomarkers and screening for cognitive frailty”. In: *Aging Clinical and Experimental Research* 29.6 (Dec. 2017), pp. 1075–1086. DOI: [10.1007/s40520-017-0741-8](https://doi.org/10.1007/s40520-017-0741-8).
- [125] Sophia Krix et al. “Towards early diagnosis of Alzheimer’s disease: advances in immune-related blood biomarkers and computational approaches”. In: *Frontiers in Immunology* 15 (Apr. 2024). DOI: [10.3389/fimmu.2024.1343900](https://doi.org/10.3389/fimmu.2024.1343900).
- [126] CR Jack et al. “Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging–Alzheimer’s Association Research Framework”. In: *JAMA Neurology* 76.10 (Oct. 2019), p. 1174. DOI: [10.1001/jamaneurol.2019.1971](https://doi.org/10.1001/jamaneurol.2019.1971).
- [127] S Martin et al. “Exploring attitudes and preferences for dementia screening in Britain: Contributions from carers and the general public”. In: *BMC Geriatrics* 15.1 (2015), pp. 1–13. DOI: [10.1186/s12877-015-0100-6](https://doi.org/10.1186/s12877-015-0100-6).
- [128] Catherine Travers, Melinda Martin-Khan, and David Lie. “Barriers and enablers of health promotion, prevention and early intervention in primary care: Evidence to inform the Australian national dementia strategy”. In: *Australasian Journal on Ageing* 28.2 (June 2009), pp. 51–57. DOI: [10.1111/j.1741-6612.2009.00359.x](https://doi.org/10.1111/j.1741-6612.2009.00359.x).
- [129] Steven Martin et al. *Attitudes and preferences towards screening for dementia: A systematic review of the literature*. June 2015. DOI: [10.1186/s12877-015-0064-6](https://doi.org/10.1186/s12877-015-0064-6).
- [130] Sarah Cullum, Rachael Simpson, and Farzana Gounder. “How is dementia portrayed in New Zealand newsprint media? Causes, effects and moral evaluation.” In: *The New Zealand medical journal* 133.1522 (Sept. 2020), pp. 18–29.
- [131] Etuini Ma’u, Sarah Cullum, and Gary Cheung. “Navigating ethnic diversity: Rethinking dementia prevention in Aotearoa New Zealand”. In: *Australian & New Zealand Journal of Psychiatry* 58.5 (May 2024), pp. 381–384. DOI: [10.1177/00048674241240586](https://doi.org/10.1177/00048674241240586).
- [132] Inga Mehrani and Perminder S. Sachdev. “The role of Memory Clinics in the assessment and management of dementia, now and into the future”. In: *Current Opinion in Psychiatry* 35.2 (Mar. 2022), pp. 118–122. DOI: [10.1097/YCO.0000000000000777](https://doi.org/10.1097/YCO.0000000000000777).

- [133] DHJ Davis et al. “Montreal Cognitive Assessment for the detection of dementia”. In: *Cochrane Database of Systematic Reviews* 2021.7 (July 2021). DOI: [10.1002/14651858.CD010775.pub3](https://doi.org/10.1002/14651858.CD010775.pub3).
- [134] M Prince et al. “The protocols for the 10/66 dementia research group population-based research programme”. In: *BMC Public Health* 7.1 (Dec. 2007), p. 165. DOI: [10.1186/1471-2458-7-165](https://doi.org/10.1186/1471-2458-7-165).
- [135] Martin J Prince. “The 10/66 dementia research group - 10 years on.” In: *Indian journal of psychiatry* 51 Suppl 1.Suppl1 (Jan. 2009), S8–S15.
- [136] Alzheimer Scotland. *Accessing post diagnostic support*. 2024. URL: <https://www.alzscot.org/pds>.
- [137] Tracy Gilmour and Post Diagnostic Practice Manager. *Facing Dementia Together - East Renfrewshire & Renfrewshire Final Report February 2009 to June 2011*. Tech. rep. 2009.
- [138] *World Alzheimer Report 2022 – Life after diagnosis: Navigating treatment, care and support*. Tech. rep.
- [139] Jennifer Kirsty Burton, Roy L Soiza, and Terence J Quinn. “Guideline summary: assessment, diagnosis, care and support for people with dementia and their carers [Scottish Intercollegiate Guidelines Network SIGN Guideline 168]”. In: *Age and Ageing* 53.7 (July 2024). DOI: [10.1093/ageing/afae147](https://doi.org/10.1093/ageing/afae147).
- [140] Kara Dassel et al. “Promoting Advance Care Planning for Persons with Dementia: Study Protocol for the LEAD (Life-Planning in Early Alzheimer’s and Other Dementias) Clinical Trial”. In: *OBM Integrative and Complementary Medicine* 08.01 (Jan. 2023), pp. 1–36. DOI: [10.21926/obm.icm.2301004](https://doi.org/10.21926/obm.icm.2301004).
- [141] P.S. Mathuranath et al. “A brief cognitive test battery to differentiate Alzheimer’s disease and frontotemporal dementia”. In: *Neurology* 55.11 (Dec. 2000), pp. 1613–1620. DOI: [10.1212/01.wnl.0000434309.85312.19](https://doi.org/10.1212/01.wnl.0000434309.85312.19).
- [142] Henry Brodaty et al. “Screening for Dementia in Primary Care: A Comparison of the GPCOG and the MMSE”. In: *Dementia and Geriatric Cognitive Disorders* 42.5-6 (2016), pp. 323–330. DOI: [10.1159/000450992](https://doi.org/10.1159/000450992).
- [143] T. Pohjasvaara et al. “Comparison of Different Clinical Criteria (DSM-III, AD-DTC, ICD-10, NINDS-AIREN, DSM-IV) for the Diagnosis of Vascular Dementia”. In: *Stroke* 31.12 (Dec. 2000), pp. 2952–2957. DOI: [10.1161/01.STR.31.12.2952](https://doi.org/10.1161/01.STR.31.12.2952).
- [144] Hude Quan et al. “Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database”. In: *Health Services Research* 43.4 (Aug. 2008), pp. 1424–1441. DOI: [10.1111/j.1475-6773.2007.00822.x](https://doi.org/10.1111/j.1475-6773.2007.00822.x).
- [145] World Health Organisation. *World Health Organisation: International Classification of Diseases (ICD) 10th Edition - Dementia*. 10th. Geneva: WHO, 1993. URL: <https://icd.who.int/browse10/2019/en#/F00.0>.
- [146] Timo Erkinjuntti. “Clinical Criteria for Vascular Dementia: The NINDS-AIREN Criteria”. In: *Dementia and Geriatric Cognitive Disorders* 5.3-4 (1994), pp. 189–192. DOI: [10.1159/000106721](https://doi.org/10.1159/000106721).

- [147] Perminder S. Sachdev et al. “Classifying neurocognitive disorders: the DSM-5 approach”. In: *Nature Reviews Neurology* 10.11 (Nov. 2014), pp. 634–642. DOI: [10.1038/nrneurol.2014.181](https://doi.org/10.1038/nrneurol.2014.181).
- [148] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)*. 5th. Arlington: American Psychiatric Association, 2013.
- [149] Bradley T. Hyman et al. “National Institute on Aging–Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease”. In: *Alzheimer’s & Dementia* 8.1 (Jan. 2012), pp. 1–13. DOI: [10.1016/j.jalz.2011.10.007](https://doi.org/10.1016/j.jalz.2011.10.007).
- [150] Pierre Molin and Kenneth Rockwood. “The New Criteria for Alzheimer’s Disease - Implications for Geriatricians*”. In: *Canadian Geriatrics Journal* 19.2 (Apr. 2016), pp. 66–73. DOI: [10.5770/cgj.19.207](https://doi.org/10.5770/cgj.19.207).
- [151] Yasumasa Yoshiyama, Virginia M Y Lee, and John Q Trojanowski. “Therapeutic strategies for tau mediated neurodegeneration”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 84.7 (July 2013), pp. 784–795. DOI: [10.1136/jnnp-2012-303144](https://doi.org/10.1136/jnnp-2012-303144).
- [152] Tim Wilkinson et al. “Identifying dementia cases with routinely collected health data: A systematic review”. In: *Alzheimer’s and Dementia* 14.8 (Aug. 2018), pp. 1038–1051. DOI: [10.1016/j.jalz.2018.02.016](https://doi.org/10.1016/j.jalz.2018.02.016).
- [153] Jacqueline C T Close. “Big data—big opportunity”. In: *Age and Ageing* 52.1 (Jan. 2023). DOI: [10.1093/ageing/afac262](https://doi.org/10.1093/ageing/afac262).
- [154] JM Brophy. “Big Data, Big Expectations, and Big Judgements”. In: *Canadian Journal of Cardiology* 38.10 (Oct. 2022), pp. 1567–1569. DOI: [10.1016/j.cjca.2022.07.001](https://doi.org/10.1016/j.cjca.2022.07.001).
- [155] Fergus N. Doubal et al. “Big data and data repurposing - using existing data to answer new questions in vascular dementia research”. In: *BMC Neurology* 17.1 (Apr. 2017). DOI: [10.1186/s12883-017-0841-2](https://doi.org/10.1186/s12883-017-0841-2).
- [156] Shawn Dolley. “Big Data’s Role in Precision Public Health”. In: *Frontiers in Public Health* 6 (Mar. 2018). DOI: [10.3389/fpubh.2018.00068](https://doi.org/10.3389/fpubh.2018.00068).
- [157] R Pastorino et al. “Benefits and challenges of Big Data in healthcare: an overview of the European initiatives”. In: *European Journal of Public Health* 29.Supplement_3 (Oct. 2019), pp. 23–27. DOI: [10.1093/eurpub/ckz168](https://doi.org/10.1093/eurpub/ckz168).
- [158] Andy Tattersall and Maria J. Grant. “Big Data – What is it and why it matters”. In: *Health Information & Libraries Journal* 33.2 (June 2016), pp. 89–91. DOI: [10.1111/hir.12147](https://doi.org/10.1111/hir.12147).
- [159] Nishita Mehta and Anil Pandit. “Concurrence of big data analytics and healthcare: A systematic review”. In: *International Journal of Medical Informatics* 114 (June 2018), pp. 57–65. DOI: [10.1016/j.ijmedinf.2018.03.013](https://doi.org/10.1016/j.ijmedinf.2018.03.013).
- [160] Andrea De Mauro, Marco Greco, and Michele Grimaldi. “What is Big Data? A Consensual Definition and a Review of Key Research Topics”. In: (2014). DOI: [10.13140/2.1.2341.5048](https://doi.org/10.13140/2.1.2341.5048).
- [161] M Aitken et al. “Consensus Statement on Public Involvement and Engagement with Data-Intensive Health Research”. In: *International Journal of Population Data Science* 4.1 (Feb. 2019). DOI: [10.23889/ijpds.v4i1.586](https://doi.org/10.23889/ijpds.v4i1.586).

- [162] A S Lunde. “The birth number concept and record linkage.” In: *American Journal of Public Health* 65.11 (Nov. 1975), pp. 1165–1169. DOI: [10.2105/AJPH.65.11.1165](https://doi.org/10.2105/AJPH.65.11.1165).
- [163] George Davey Smith. “Commentary: Behind the Broad Street pump: aetiology, epidemiology and prevention of cholera in mid-19th century Britain”. In: *International Journal of Epidemiology* 31.5 (Oct. 2002), pp. 920–932. DOI: [10.1093/ije/31.5.920](https://doi.org/10.1093/ije/31.5.920).
- [164] MA Beyer and D Laney. *The importance of "Big Data": A definition*. Gartner. 2012. URL: <https://www.gartner.com/en/documents/2057415>.
- [165] Jakub Mlacki. *7 V's of big data - what do you need to know?* URL: <https://dsstream.com/7-vs-of-big-data-what-are-they-and-why-are-they-so-important/>.
- [166] Continuous(TM) Care for health. *What is the difference between EMR, EHR and PHR?* 2020. URL: <https://www.continuouscare.io/blog/difference-between-emr-ehr-phr/>.
- [167] Josh Keith, Fiona Grimm, and Adam Steventon. *How better use of data can help address key challenges facing the NHS*. Tech. rep. 2022.
- [168] World Health Organisation. *Global strategy on digital health 2020-2025*. 2021. ISBN: 9789240020924. URL: <http://apps.who.int/bookorders..>
- [169] Marcus Richards. “The Power of Birth Cohorts to Study Risk Factors for Cognitive Impairment”. In: *Current Neurology and Neuroscience Reports* 22.12 (Dec. 2022), pp. 847–854. DOI: [10.1007/s11910-022-01244-0](https://doi.org/10.1007/s11910-022-01244-0).
- [170] Yuan Lin et al. “The continuing evolution of birth cohort studies: achievements and challenges”. In: *Biology of Reproduction* 107.1 (July 2022), pp. 358–367. DOI: [10.1093/biolre/iaoac117](https://doi.org/10.1093/biolre/iaoac117).
- [171] Silvia S. Klokgieters et al. “Changes in the role of explanatory factors for socioeconomic inequalities in physical performance: a comparative study of three birth cohorts”. In: *International Journal for Equity in Health* 20.1 (Dec. 2021), p. 252. DOI: [10.1186/s12939-021-01592-2](https://doi.org/10.1186/s12939-021-01592-2).
- [172] Maria Torres Toda et al. “Exposure to natural environments during pregnancy and birth outcomes in 11 European birth cohorts”. In: *Environment International* 170 (Dec. 2022), p. 107648. DOI: [10.1016/j.envint.2022.107648](https://doi.org/10.1016/j.envint.2022.107648).
- [173] L. Li et al. “Maximising the potential of Chinese birth cohort studies: a systematic review of mother–baby cohorts in mainland China”. In: *Public Health* 227 (Feb. 2024), pp. 119–130. DOI: [10.1016/j.puhe.2023.11.035](https://doi.org/10.1016/j.puhe.2023.11.035).
- [174] Marcus Pembrey. “Genetic epidemiology: some special contributions of birth cohorts”. In: *Paediatric and Perinatal Epidemiology* 18.1 (Jan. 2004), pp. 3–7. DOI: [10.1111/j.1365-3016.2004.00530.x](https://doi.org/10.1111/j.1365-3016.2004.00530.x).
- [175] C Canova and A Cantarutti. “Population-Based Birth Cohort Studies in Epidemiology”. In: *International Journal of Environmental Research and Public Health* 17.15 (July 2020), p. 5276. DOI: [10.3390/ijerph17155276](https://doi.org/10.3390/ijerph17155276).
- [176] Sarah-Naomi James et al. “Using a birth cohort to study brain health and pre-clinical dementia: recruitment and participation rates in Insight 46”. In: *BMC Research Notes* 11.1 (Dec. 2018), p. 885. DOI: [10.1186/s13104-018-3995-0](https://doi.org/10.1186/s13104-018-3995-0).

- [177] Christopher A. Lane et al. “Study protocol: Insight 46 – a neuroscience sub-study of the MRC National Survey of Health and Development”. In: *BMC Neurology* 17.1 (Dec. 2017), p. 75. DOI: [10.1186/s12883-017-0846-x](https://doi.org/10.1186/s12883-017-0846-x).
- [178] M. Stafford et al. “Using a birth cohort to study ageing: representativeness and response rates in the National Survey of Health and Development”. In: *European Journal of Ageing* 10.2 (June 2013), pp. 145–157. DOI: [10.1007/s10433-013-0258-8](https://doi.org/10.1007/s10433-013-0258-8).
- [179] Beate H. Danielsen et al. “Linked birth cohort files for perinatal health research: California as a model for methodology and implementation”. In: *Annals of Epidemiology* 79 (Mar. 2023), pp. 10–18. DOI: [10.1016/j.annepidem.2022.12.014](https://doi.org/10.1016/j.annepidem.2022.12.014).
- [180] Catherine H Stewart, Ruth Dundas, and Alastair H Leyland. “The Scottish school leavers cohort: linkage of education data to routinely collected records for mortality, hospital discharge and offspring birth characteristics”. In: *BMJ Open* 7.7 (July 2017), e015027. DOI: [10.1136/bmjopen-2016-015027](https://doi.org/10.1136/bmjopen-2016-015027).
- [181] KS Tingay et al. “Record linkage to enhance consented cohort and routinely collected health data from a UK birth cohort”. In: *International Journal of Population Data Science* 4.1 (Apr. 2019). DOI: [10.23889/ijpds.v4i1.579](https://doi.org/10.23889/ijpds.v4i1.579).
- [182] Gill Livingston et al. “Dementia prevention, intervention, and care: 2020 report of the Lancet Commission”. In: *The Lancet* 396.10248 (Aug. 2020), pp. 413–446. DOI: [10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6).
- [183] Reisa A. Sperling et al. “Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease”. In: *Alzheimer’s & Dementia* 7.3 (May 2011), pp. 280–292. DOI: [10.1016/j.jalz.2011.03.003](https://doi.org/10.1016/j.jalz.2011.03.003).
- [184] CR Jack et al. “Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade”. In: *The Lancet Neurology* 9.1 (Jan. 2010), pp. 119–128. DOI: [10.1016/S1474-4422\(09\)70299-6](https://doi.org/10.1016/S1474-4422(09)70299-6).
- [185] Guangning Wang et al. “Game-based brain training for improving cognitive function in community-dwelling older adults: A systematic review and meta-regression”. In: *Archives of Gerontology and Geriatrics* 92 (Jan. 2021), p. 104260. DOI: [10.1016/j.archger.2020.104260](https://doi.org/10.1016/j.archger.2020.104260).
- [186] Fumihiko Taya et al. “Brain enhancement through cognitive training: a new insight from brain connectome”. In: *Frontiers in Systems Neuroscience* 9 (Apr. 2015). DOI: [10.3389/fnsys.2015.00044](https://doi.org/10.3389/fnsys.2015.00044).
- [187] Zoltan Ungvari et al. “The multifaceted benefits of walking for healthy aging: from Blue Zones to molecular mechanisms”. In: *GeroScience* 45.6 (July 2023), pp. 3211–3239. DOI: [10.1007/s11357-023-00873-8](https://doi.org/10.1007/s11357-023-00873-8).
- [188] LM Besser et al. “The Healthy Brain Initiative (HBI): A prospective cohort study protocol”. In: *PLOS ONE* 18.10 (Oct. 2023), e0293634. DOI: [10.1371/journal.pone.0293634](https://doi.org/10.1371/journal.pone.0293634).
- [189] World Health Organisation. *Progress report on the United Nations Decade of Healthy Ageing, 2021-2023*. Tech. rep. Geneva, 2023.

- [190] David A. Snowdon. “Healthy Aging and Dementia: Findings from the Nun Study”. In: *Annals of Internal Medicine* 139.5_Part_2 (Sept. 2003), p. 450. DOI: [10.7326/0003-4819-139-5_Part_2-200309021-00014](https://doi.org/10.7326/0003-4819-139-5_Part_2-200309021-00014).
- [191] Y Ben-Shlomo and D Kuh. “A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives.” In: *International journal of epidemiology* 31.2 (Apr. 2002), pp. 285–293. DOI: [10.1093/ije/31.2.285](https://doi.org/10.1093/ije/31.2.285).
- [192] DJP Barker. “EDITORIAL: The developmental origins of adult disease”. In: *European Journal of Epidemiology* 18.8 (Aug. 2002), pp. 733–736. DOI: [10.1023/A:1025388901248](https://doi.org/10.1023/A:1025388901248).
- [193] R. Bonita, R. Beaglehole, and T. Kjellström. *Basic epidemiology*. Geneva: World Health Organisation, 2006. URL: https://drive.uqu.edu.sa/_/maatia/files/Basic%20epidemiology.pdf.
- [194] Aline Marileen Wieggersma et al. “Do prenatal factors shape the risk for dementia?: A systematic review of the epidemiological evidence for the prenatal origins of dementia”. In: *Social Psychiatry and Psychiatric Epidemiology* (Apr. 2023). DOI: [10.1007/s00127-023-02471-7](https://doi.org/10.1007/s00127-023-02471-7).
- [195] Susan D. Shenkin, Ian J. Deary, and John M. Starr. “Birth Parameters and Cognitive Ability in Older Age: A Follow-Up Study of People Born 1921–1926”. In: *Gerontology* 55.1 (2009), pp. 92–98. DOI: [10.1159/000163444](https://doi.org/10.1159/000163444).
- [196] Susan D. Shenkin, John M. Starr, and Ian J. Deary. “Birth Weight and Cognitive Ability in Childhood: A Systematic Review.” In: *Psychological Bulletin* 130.6 (2004), pp. 989–1013. DOI: [10.1037/0033-2909.130.6.989](https://doi.org/10.1037/0033-2909.130.6.989).
- [197] Luigi Yuri Di Marco et al. “Modifiable Lifestyle Factors in Dementia: A Systematic Review of Longitudinal Observational Cohort Studies”. In: *Journal of Alzheimer’s Disease* 42.1 (Aug. 2014), pp. 119–135. DOI: [10.3233/JAD-132225](https://doi.org/10.3233/JAD-132225).
- [198] Gill Livingston et al. “Dementia prevention, intervention, and care”. In: *The Lancet* 390.10113 (Dec. 2017), pp. 2673–2734. DOI: [10.1016/S0140-6736\(17\)31363-6](https://doi.org/10.1016/S0140-6736(17)31363-6).
- [199] Mohammad Ali Mansournia and Douglas G Altman. “Population attributable fraction”. In: *BMJ* (Feb. 2018), k757. DOI: [10.1136/bmj.k757](https://doi.org/10.1136/bmj.k757).
- [200] World Health Organisation. *Brain health*. 2024. URL: https://www.who.int/health-topics/brain-health#tab=tab_1.
- [201] World Health Organisation. *Optimizing brain health across the life course*. Tech. rep. 2022. URL: <http://apps.who.int/bookorders..>
- [202] World Health Organisation. *Intersectoral global action plan on epilepsy and other neurological disorders*. Tech. rep. 2023.
- [203] Sarah Lenz Lock et al. “Equity across the life course matters for brain health”. In: *Nature Aging* 3.5 (May 2023), pp. 466–468. DOI: [10.1038/s43587-023-00413-1](https://doi.org/10.1038/s43587-023-00413-1).
- [204] Global Council on Brain Health. *How to Sustain Healthy Behaviors: Applying Lessons of Public Health and Science to Drive Change*. Tech. rep. Washington, DC: Global Council on Brain Health, Mar. 2022. DOI: [10.26419/pia.00106.001](https://doi.org/10.26419/pia.00106.001).

- [205] John Zhou, Marc Benoit, and Md Golam Sharoar. “Recent advances in pre-clinical diagnosis of Alzheimer’s disease”. In: *Metabolic Brain Disease* 37.6 (Aug. 2022), pp. 1703–1725. DOI: [10.1007/s11011-021-00733-4](https://doi.org/10.1007/s11011-021-00733-4).
- [206] Lisa Vermunt et al. “Duration of preclinical, prodromal, and dementia stages of Alzheimer’s disease in relation to age, sex, and <i>APOE</i> genotype”. In: *Alzheimer’s & Dementia* 15.7 (July 2019), pp. 888–898. DOI: [10.1016/j.jalz.2019.04.001](https://doi.org/10.1016/j.jalz.2019.04.001).
- [207] Wei Hu et al. “Associations and Mediating Pathways Between Childhood Adversity and Risk of Dementia: A Cohort Study in the UK Biobank”. In: *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* 79.8 (Aug. 2024). DOI: [10.1093/gerona/glae121](https://doi.org/10.1093/gerona/glae121).
- [208] Kaarina Korhonen et al. “Childhood socio-economic circumstances and dementia: prospective register-based cohort study of adulthood socio-economic and cardiovascular health mediators”. In: *International Journal of Epidemiology* 52.2 (Apr. 2023), pp. 523–535. DOI: [10.1093/ije/dyac205](https://doi.org/10.1093/ije/dyac205).
- [209] Dinuli Nilaweera et al. “The association between adverse childhood events and later-life cognitive function and dementia risk”. In: *Journal of Affective Disorders* 304 (May 2022), pp. 128–132. DOI: [10.1016/j.jad.2022.02.062](https://doi.org/10.1016/j.jad.2022.02.062).
- [210] Yue-Ting Deng et al. “Association of life course adiposity with risk of incident dementia: a prospective cohort study of 322,336 participants”. In: *Molecular Psychiatry* 27.8 (Aug. 2022), pp. 3385–3395. DOI: [10.1038/s41380-022-01604-9](https://doi.org/10.1038/s41380-022-01604-9).
- [211] X-J Wang et al. “Early-Life Risk Factors for Dementia and Cognitive Impairment in Later Life: A Systematic Review and Meta-Analysis”. In: *Journal of Alzheimer’s Disease* 67.1 (Jan. 2019), pp. 221–229. DOI: [10.3233/JAD-180856](https://doi.org/10.3233/JAD-180856).
- [212] S. Zhou and K. Wang. “Childhood Secondhand Smoke Exposure and Risk of Dementia, Alzheimer’s Disease and Stroke in Adulthood: A Prospective Cohort Study”. In: *Journal of Prevention of Alzheimer’s Disease* (2021), pp. 1–6. DOI: [10.14283/jpad.2021.10](https://doi.org/10.14283/jpad.2021.10).
- [213] Caroline A. McHutchison et al. “Education, Socioeconomic Status, and Intelligence in Childhood and Stroke Risk in Later Life”. In: *Epidemiology* 28.4 (July 2017), pp. 608–618. DOI: [10.1097/EDE.0000000000000675](https://doi.org/10.1097/EDE.0000000000000675).
- [214] Amy R Borenstein, Cathleen I Copenhaver, and James A Mortimer. “Early-Life Risk Factors for Alzheimer Disease”. In: *Alzheimer Disease & Associated Disorders* 20.1 (Jan. 2006), pp. 63–72. DOI: [10.1097/01.wad.0000201854.62116.d7](https://doi.org/10.1097/01.wad.0000201854.62116.d7).
- [215] Y Tani, T Fujiwara, and K Kondo. “Association Between Adverse Childhood Experiences and Dementia in Older Japanese Adults”. In: *JAMA Network Open* 3.2 (Feb. 2020), e1920740. DOI: [10.1001/jamanetworkopen.2019.20740](https://doi.org/10.1001/jamanetworkopen.2019.20740).
- [216] Emilia Severs et al. “Traumatic life events and risk for dementia: a systematic review and meta-analysis”. In: *BMC Geriatrics* 23.1 (Sept. 2023), p. 587. DOI: [10.1186/s12877-023-04287-1](https://doi.org/10.1186/s12877-023-04287-1).
- [217] Brian McGurn, Ian J. Deary, and John M. Starr. “Childhood cognitive ability and risk of late-onset Alzheimer and vascular dementia”. In: *Neurology* 71.14 (Sept. 2008), pp. 1051–1056. DOI: [10.1212/01.wnl.0000319692.20283.10](https://doi.org/10.1212/01.wnl.0000319692.20283.10).

- [218] CS Widom et al. “Childhood Maltreatment and Dementia Risk Factors in Midlife: A Prospective Investigation”. In: *Current Alzheimer Research* 20.9 (Sept. 2023), pp. 636–647. DOI: [10.2174/0115672050281539231222071355](https://doi.org/10.2174/0115672050281539231222071355).
- [219] Havisha Pedamallu et al. “Life Course Cardiovascular Health: Risk Factors, Outcomes, and Interventions”. In: *Circulation Research* 132.12 (June 2023), pp. 1570–1583. DOI: [10.1161/CIRCRESAHA.123.321998](https://doi.org/10.1161/CIRCRESAHA.123.321998).
- [220] GD Batty, IJ. Deary, and LS Gottfredson. “Premorbid (early life) IQ and Later Mortality Risk: Systematic Review”. In: *Annals of Epidemiology* 17.4 (Apr. 2007), pp. 278–288. DOI: [10.1016/j.annepidem.2006.07.010](https://doi.org/10.1016/j.annepidem.2006.07.010).
- [221] M Bloomberg et al. “Healthy lifestyle and cognitive decline in middle-aged and older adults residing in 14 European countries”. In: *Nature Communications* 15.1 (June 2024), p. 5003. DOI: [10.1038/s41467-024-49262-5](https://doi.org/10.1038/s41467-024-49262-5).
- [222] Craig W Ritchie et al. “The PREVENT dementia programme: baseline demographic, lifestyle, imaging and cognitive data from a midlife cohort study investigating risk factors for dementia”. In: *Brain Communications* 6.3 (May 2024). DOI: [10.1093/braincomms/fcae189](https://doi.org/10.1093/braincomms/fcae189).
- [223] Dennis Chan et al. “Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation, and late-life activities”. In: *Neurobiology of Aging* 70 (Oct. 2018), pp. 180–183. DOI: [10.1016/j.neurobiolaging.2018.06.012](https://doi.org/10.1016/j.neurobiolaging.2018.06.012).
- [224] Aaron Reuben et al. “Improving risk indexes for Alzheimer’s disease and related dementias for use in midlife”. In: *Brain Communications* 4.5 (Sept. 2022). DOI: [10.1093/braincomms/fcac223](https://doi.org/10.1093/braincomms/fcac223).
- [225] Justin M. Long and David M. Holtzman. “Alzheimer Disease: An Update on Pathobiology and Treatment Strategies”. In: *Cell* 179.2 (Oct. 2019), pp. 312–339. DOI: [10.1016/j.cell.2019.09.001](https://doi.org/10.1016/j.cell.2019.09.001).
- [226] Barry Bogin and Maria Inês Varela-Silva. “Leg Length, Body Proportion, and Health: A Review with a Note on Beauty”. In: *International Journal of Environmental Research and Public Health* 7.3 (Mar. 2010), pp. 1047–1075. DOI: [10.3390/ijerph7031047](https://doi.org/10.3390/ijerph7031047).
- [227] D. J. Gunnell et al. “Childhood leg length and adult mortality: follow up of the Carnegie (Boyd Orr) Survey of Diet and Health in Pre-war Britain”. In: *Journal of Epidemiology & Community Health* 52.3 (Mar. 1998), pp. 142–152. DOI: [10.1136/jech.52.3.142](https://doi.org/10.1136/jech.52.3.142).
- [228] Rachel Denholm, Chris Power, and Leah Li. “Adverse childhood experiences and child-to-adult height trajectories in the 1958 British birth cohort”. In: *International Journal of Epidemiology* 42.5 (Oct. 2013), pp. 1399–1409. DOI: [10.1093/ije/dyt169](https://doi.org/10.1093/ije/dyt169).
- [229] Chao Wang et al. “Association between Exposure to the Chinese Famine in Different Stages of Early Life and Decline in Cognitive Functioning in Adulthood”. In: *Frontiers in Behavioral Neuroscience* 10 (July 2016). DOI: [10.3389/fnbeh.2016.00146](https://doi.org/10.3389/fnbeh.2016.00146).
- [230] Yadollah Momtaz et al. “Does food insufficiency in childhood contribute to dementia in later life?” In: *Clinical Interventions in Aging* (Dec. 2014), p. 49. DOI: [10.2147/CIA.S69220](https://doi.org/10.2147/CIA.S69220).

- [231] J-M Kim. “Limb length and dementia in an older Korean population”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 74.4 (Apr. 2003), pp. 427–432. DOI: [10.1136/jnnp.74.4.427](https://doi.org/10.1136/jnnp.74.4.427).
- [232] Jae-Min Kim et al. “Associations between head circumference, leg length and dementia in a Korean population”. In: *International Journal of Geriatric Psychiatry* 23.1 (Jan. 2008), pp. 41–48. DOI: [10.1002/gps.1833](https://doi.org/10.1002/gps.1833).
- [233] Z Mak, J-M Kim, and R Stewart. “Leg length, cognitive impairment and cognitive decline in an African-Caribbean population”. In: *International Journal of Geriatric Psychiatry* 21.3 (Mar. 2006), pp. 266–272. DOI: [10.1002/gps.1458](https://doi.org/10.1002/gps.1458).
- [234] William S. Kremen et al. “Influence of young adult cognitive ability and additional education on later-life cognition”. In: *Proceedings of the National Academy of Sciences* 116.6 (Feb. 2019), pp. 2021–2026. DOI: [10.1073/pnas.1811537116](https://doi.org/10.1073/pnas.1811537116).
- [235] Kathleen S. Hall et al. “Low education and childhood rural residence”. In: *Neurology* 54.1 (Jan. 2000), pp. 95–95. DOI: [10.1212/WNL.54.1.95](https://doi.org/10.1212/WNL.54.1.95).
- [236] Israel Contador et al. “Childhood and Adulthood Rural Residence Increases the Risk of Dementia: NEDICES Study”. In: *Current Alzheimer Research* 12.4 (Apr. 2015), pp. 350–357. DOI: [10.2174/1567205012666150324181327](https://doi.org/10.2174/1567205012666150324181327).
- [237] Yaakov Stern et al. “Brain reserve, cognitive reserve, compensation, and maintenance: operationalization, validity, and mechanisms of cognitive resilience”. In: *Neurobiology of Aging* 83 (Nov. 2019), pp. 124–129. DOI: [10.1016/j.neurobiolaging.2019.03.022](https://doi.org/10.1016/j.neurobiolaging.2019.03.022).
- [238] Yaakov Stern and Daniel Barulli. “Cognitive reserve”. In: 2019, pp. 181–190. DOI: [10.1016/B978-0-12-804766-8.00011-X](https://doi.org/10.1016/B978-0-12-804766-8.00011-X).
- [239] Melis Anatürk et al. “Prediction of brain age and cognitive age: Quantifying brain and cognitive maintenance in aging”. In: *Human Brain Mapping* 42.6 (Apr. 2021), pp. 1626–1640. DOI: [10.1002/hbm.25316](https://doi.org/10.1002/hbm.25316).
- [240] Claudia Schwarz et al. “Reserve, resilience and maintenance of episodic memory and other cognitive functions in aging”. In: *Neurobiology of Aging* 140 (Aug. 2024), pp. 60–69. DOI: [10.1016/j.neurobiolaging.2024.04.011](https://doi.org/10.1016/j.neurobiolaging.2024.04.011).
- [241] Global Council on Brain Health. *Engage Your Brain: GCBH Recommendations on Cognitively Stimulating Activities*. Tech. rep. Washington, DC: Global Council on Brain Health, July 2017. DOI: [10.26419/pia.00001.001](https://doi.org/10.26419/pia.00001.001).
- [242] RT Staff et al. “Intellectual engagement and cognitive ability in later life (the “use it or lose it” conjecture): longitudinal, prospective study”. In: *BMJ* (Dec. 2018), k4925. DOI: [10.1136/bmj.k4925](https://doi.org/10.1136/bmj.k4925).
- [243] Allen T. C. Lee et al. “Association of Daily Intellectual Activities With Lower Risk of Incident Dementia Among Older Chinese Adults”. In: *JAMA Psychiatry* 75.7 (July 2018), p. 697. DOI: [10.1001/jamapsychiatry.2018.0657](https://doi.org/10.1001/jamapsychiatry.2018.0657).
- [244] Stephanie J.B. Vos et al. “Modifiable Risk Factors for Prevention of Dementia in Midlife, Late Life and the Oldest-Old: Validation of the LIBRA Index”. In: *Journal of Alzheimer’s Disease* 58.2 (May 2017), pp. 537–547. DOI: [10.3233/JAD-161208](https://doi.org/10.3233/JAD-161208).

- [245] Kaarin J. Anstey, Nicolas Cherbuin, and Pushpani M Herath. “Development of a New Method for Assessing Global Risk of Alzheimer’s Disease for Use in Population Health Approaches to Prevention”. In: *Prevention Science* 14.4 (Aug. 2013), pp. 411–421. DOI: [10.1007/s11121-012-0313-2](https://doi.org/10.1007/s11121-012-0313-2).
- [246] Miia Kivipelto et al. “Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study”. In: *The Lancet Neurology* 5.9 (Sept. 2006), pp. 735–741. DOI: [10.1016/S1474-4422\(06\)70537-3](https://doi.org/10.1016/S1474-4422(06)70537-3).
- [247] S Licher et al. “Development and Validation of a Dementia Risk Prediction Model in the General Population: An Analysis of Three Longitudinal Studies”. In: *American Journal of Psychiatry* 176.7 (July 2019), pp. 543–551. DOI: [10.1176/appi.ajp.2018.18050566](https://doi.org/10.1176/appi.ajp.2018.18050566).
- [248] Eugene Y. H. Tang et al. “Current Developments in Dementia Risk Prediction Modelling: An Updated Systematic Review”. In: *PLOS ONE* 10.9 (Sept. 2015), e0136181. DOI: [10.1371/journal.pone.0136181](https://doi.org/10.1371/journal.pone.0136181).
- [249] Aurore Fayosse et al. “Risk prediction models for dementia: role of age and cardiometabolic risk factors”. In: *BMC Medicine* 18.1 (Dec. 2020), p. 107. DOI: [10.1186/s12916-020-01578-x](https://doi.org/10.1186/s12916-020-01578-x).
- [250] Lingling Wang et al. “Construction of a risk prediction model for Alzheimer’s disease in the elderly population”. In: *BMC Neurology* 21.1 (Dec. 2021), p. 271. DOI: [10.1186/s12883-021-02276-8](https://doi.org/10.1186/s12883-021-02276-8).
- [251] Silvan Licher et al. “External validation of four dementia prediction models for use in the general community-dwelling population: a comparative analysis from the Rotterdam Study”. In: *European Journal of Epidemiology* 33.7 (July 2018), pp. 645–655. DOI: [10.1007/s10654-018-0403-y](https://doi.org/10.1007/s10654-018-0403-y).
- [252] OY Bello-Chavolla, CA Aguilar-Salinas, and JA Avila-Funes. “The type 2 diabetes-specific dementia risk score (DSDRS) is associated with frailty, cognitive and functional status amongst Mexican community-dwelling older adults”. In: *BMC Geriatrics* 20.1 (Dec. 2020), p. 363. DOI: [10.1186/s12877-020-01776-5](https://doi.org/10.1186/s12877-020-01776-5).
- [253] J Brain et al. “What’s new in dementia risk prediction modelling? An updated systematic review”. In: *Dementia and Geriatric Cognitive Disorders Extra* (June 2024). DOI: [10.1159/000539744](https://doi.org/10.1159/000539744).
- [254] CW Ritchie and G Muniz-Terrera. “Models for dementia risk prediction: so much activity brings a need for coordination and clarity”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 90.4 (Apr. 2019), pp. 372–372. DOI: [10.1136/jnnp-2018-318944](https://doi.org/10.1136/jnnp-2018-318944).
- [255] K Anstey and N Cherbuin. *The Australian National University Alzheimer’s Disease Risk Index (ANU-ADRI)*. URL: <https://anuadri.anu.edu.au/>.
- [256] Robin Kearns and Damian Collins. “Health Geography”. In: *A Companion to Health and Medical Geography*. Wiley, Dec. 2009, pp. 13–32. DOI: [10.1002/9781444314762.ch2](https://doi.org/10.1002/9781444314762.ch2).
- [257] T Brown, S McLafferty, and G Moon. *Introduction to Health and Medical Geography*. Wiley online, Blackwell Publishing Ltd., Dec. 2009, pp. 1–11. ISBN: 9781444314762(online). DOI: [10.1002/9781444314762.ch1](https://doi.org/10.1002/9781444314762.ch1).

- [258] Jonathan D. Mayer. “Medical Geography”. In: *A Companion to Health and Medical Geography*. Wiley, Dec. 2009, pp. 33–54. DOI: [10.1002/9781444314762.ch3](https://doi.org/10.1002/9781444314762.ch3).
- [259] Sarah Curtis, Mylène Riva, and Mark Rosenberg. “Health Geography and Public Health”. In: *A Companion to Health and Medical Geography*. Wiley, Dec. 2009, pp. 323–345. DOI: [10.1002/9781444314762.ch18](https://doi.org/10.1002/9781444314762.ch18).
- [260] G Moon. “Health Geography”. In: *International Encyclopedia of Human Geography, Second Edition*. Elsevier, Jan. 2019, pp. 315–321. ISBN: 9780081022955. DOI: [10.1016/B978-0-08-102295-5.10388-9](https://doi.org/10.1016/B978-0-08-102295-5.10388-9).
- [261] AC Gatrell and SJ Elliott. *Geographies of health : an introduction*. Chichester, England: Wiley Blackwell, 2015.
- [262] P Boyle et al. *The Geography of Health Inequalities in the Developed World: Views from Britain and North America*. Burlington, VT: Ashgate Publishing, 2004.
- [263] Lewis O. J. Killin et al. “Environmental risk factors for dementia: a systematic review”. In: *BMC Geriatrics* 16.1 (Dec. 2016), p. 175. DOI: [10.1186/s12877-016-0342-y](https://doi.org/10.1186/s12877-016-0342-y).
- [264] G Dean and JF Kurtzke. “On the Risk of Multiple Sclerosis According to Age at Immigration to South Africa”. In: *BMJ* 3.5777 (Sept. 1971), pp. 725–729. DOI: [10.1136/bmj.3.5777.725](https://doi.org/10.1136/bmj.3.5777.725).
- [265] S. R. Hammond. “The age-range of risk of developing multiple sclerosis: Evidence from a migrant population in Australia”. In: *Brain* 123.5 (May 2000), pp. 968–974. DOI: [10.1093/brain/123.5.968](https://doi.org/10.1093/brain/123.5.968).
- [266] M. Powell Lawton and Lucille Nahemow. “Ecology and the aging process.” In: *The psychology of adult development and aging*. Washington: American Psychological Association, 1973, pp. 619–674. DOI: [10.1037/10044-020](https://doi.org/10.1037/10044-020).
- [267] M. P. Lawton and B. Simon. “The Ecology of Social Relationships in Housing for the Elderly”. In: *The Gerontologist* 8.2 (June 1968), pp. 108–115. DOI: [10.1093/geront/8.2.108](https://doi.org/10.1093/geront/8.2.108).
- [268] M Cassarino and A Setti. “Environment as ‘Brain Training’: A review of geographical and physical environmental influences on cognitive ageing”. In: *Ageing Research Reviews* 23 (Sept. 2015), pp. 167–182. DOI: [10.1016/j.arr.2015.06.003](https://doi.org/10.1016/j.arr.2015.06.003).
- [269] Hélène T Cronjé et al. “Leveraging the Urban–Rural Divide for Epigenetic Research”. In: *Epigenomics* 12.12 (June 2020), pp. 1071–1081. DOI: [10.2217/epi-2020-0049](https://doi.org/10.2217/epi-2020-0049).
- [270] Tamas Pandics et al. “Exposome and unhealthy aging: environmental drivers from air pollution to occupational exposures”. In: *GeroScience* 45.6 (Sept. 2023), pp. 3381–3408. DOI: [10.1007/s11357-023-00913-3](https://doi.org/10.1007/s11357-023-00913-3).
- [271] CP Wild. “Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology”. In: *Cancer Epidemiology, Biomarkers & Prevention* 14.8 (Aug. 2005), pp. 1847–1850. DOI: [10.1158/1055-9965.EPI-05-0456](https://doi.org/10.1158/1055-9965.EPI-05-0456).
- [272] Christopher Paul Wild. “The exposome: from concept to utility”. In: *International Journal of Epidemiology* 41.1 (Feb. 2012), pp. 24–32. DOI: [10.1093/ije/dyr236](https://doi.org/10.1093/ije/dyr236).

- [273] CP Wild, A Scalbert, and Z Herceg. “Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk”. In: *Environmental and Molecular Mutagenesis* 54.7 (Aug. 2013), pp. 480–499. DOI: [10.1002/em.21777](https://doi.org/10.1002/em.21777).
- [274] Martine Vrijheid. “The exposome: a new paradigm to study the impact of environment on health”. In: *Thorax* 69.9 (Sept. 2014), pp. 876–878. DOI: [10.1136/thoraxjnl-2013-204949](https://doi.org/10.1136/thoraxjnl-2013-204949).
- [275] MM Niedzwiecki and GW Miller. “The Exposome Paradigm in Human Health: Lessons from the Emory Exposome Summer Course”. In: *Environmental Health Perspectives* 125.6 (June 2017). DOI: [10.1289/EHP1712](https://doi.org/10.1289/EHP1712).
- [276] A Guillien et al. “The Exposome Approach to Decipher the Role of Multiple Environmental and Lifestyle Determinants in Asthma”. In: *International Journal of Environmental Research and Public Health* 18.3 (Jan. 2021), p. 1138. DOI: [10.3390/ijerph18031138](https://doi.org/10.3390/ijerph18031138).
- [277] United Nations. *United Nations Goal 11: Sustainable cities and communities*. URL: <https://unstats.un.org/sdgs/report/2023/goal-11/>.
- [278] Internet Geography. *Urbanisation*. URL: <http://www.bennett.karoo.net/topics/urban.html#urb>.
- [279] European Environment Agency (EEA). *Urbanisation*. URL: <https://www.eea.europa.eu/help/glossary/eea-glossary/urbanisation>.
- [280] Mairead M. Bartley et al. “Social Determinants of Health among Older Adults with Dementia in Urban and Rural Areas”. In: *The Journal of Prevention of Alzheimer’s Disease* 10.4 (Aug. 2023), pp. 895–902. DOI: [10.1007/s42414-023-0002-2](https://doi.org/10.1007/s42414-023-0002-2).
- [281] E Harries. “Research to Practice Paper: Social Isolation and its Relationship to the Urban Environment”. PhD thesis. McGill University, pp. 1–13.
- [282] *The Ten and a Half Myths that may Distort the urban Policies of Governments and International Agencies*. Tech. rep.
- [283] Hannah Badland et al. “Urban liveability: Emerging lessons from Australia for exploring the potential for indicators to measure the social determinants of health”. In: *Social Science & Medicine* 111 (June 2014), pp. 64–73. DOI: [10.1016/j.socscimed.2014.04.003](https://doi.org/10.1016/j.socscimed.2014.04.003).
- [284] X Zhang et al. “A systematic review of the anxiety-alleviation benefits of exposure to the natural environment”. In: *Reviews on Environmental Health* 38.2 (June 2023), pp. 281–293. DOI: [10.1515/reveh-2021-0157](https://doi.org/10.1515/reveh-2021-0157).
- [285] TL Milfont and PW Schultz. “Culture and the natural environment”. In: *Current Opinion in Psychology* 8 (Apr. 2016), pp. 194–199. DOI: [10.1016/j.copsyc.2015.09.009](https://doi.org/10.1016/j.copsyc.2015.09.009).
- [286] LA Paul et al. “Urban green space and the risks of dementia and stroke”. In: *Environmental Research* 186 (July 2020), p. 109520. DOI: [10.1016/j.envres.2020.109520](https://doi.org/10.1016/j.envres.2020.109520).
- [287] Alex de la Iglesia Martinez and S.M. Labib. “Demystifying normalized difference vegetation index (NDVI) for greenness exposure assessments and policy interventions in urban greening”. In: *Environmental Research* 220 (Mar. 2023), p. 115155. DOI: [10.1016/j.envres.2022.115155](https://doi.org/10.1016/j.envres.2022.115155).

- [288] E. S. PAYKEL et al. “Urban–rural mental health differences in Great Britain: findings from the National Morbidity Survey”. In: *Psychological Medicine* 30.2 (Mar. 2000), pp. 269–280. DOI: [10.1017/S003329179900183X](https://doi.org/10.1017/S003329179900183X).
- [289] Fiona K Judd et al. “High Prevalence Disorders in Urban and Rural Communities”. In: *Australian & New Zealand Journal of Psychiatry* 36.1 (Feb. 2002), pp. 104–113. DOI: [10.1046/j.1440-1614.2002.00986.x](https://doi.org/10.1046/j.1440-1614.2002.00986.x).
- [290] Sushmitha Kasturi et al. “Prevalence Rates of Depression and Anxiety among Young Rural and Urban Australians: A Systematic Review and Meta-Analysis”. In: *International Journal of Environmental Research and Public Health* 20.1 (Jan. 2023), p. 800. DOI: [10.3390/ijerph20010800](https://doi.org/10.3390/ijerph20010800).
- [291] Sarah Romans, Marsha Cohen, and Tonia Forte. “Rates of depression and anxiety in urban and rural Canada”. In: *Social Psychiatry and Psychiatric Epidemiology* 46.7 (July 2011), pp. 567–575. DOI: [10.1007/s00127-010-0222-2](https://doi.org/10.1007/s00127-010-0222-2).
- [292] Levi N. Bonnell et al. “Urban–Rural Differences in Mental and Physical Health among Primary Care Patients with Multiple Chronic Conditions: A Secondary Analysis from a Randomized Clinical Trial”. In: *International Journal of Environmental Research and Public Health* 19.23 (Nov. 2022), p. 15580. DOI: [10.3390/ijerph192315580](https://doi.org/10.3390/ijerph192315580).
- [293] Karen McKenzie, Aja Murray, and Tom Booth. “Do urban environments increase the risk of anxiety, depression and psychosis? An epidemiological study”. In: *Journal of Affective Disorders* 150.3 (Sept. 2013), pp. 1019–1024. DOI: [10.1016/j.jad.2013.05.032](https://doi.org/10.1016/j.jad.2013.05.032).
- [294] Ahmad Bolouki. “Neurobiological effects of urban built and natural environment on mental health: systematic review”. In: *Reviews on Environmental Health* 38.1 (Mar. 2023), pp. 169–179. DOI: [10.1515/reveh-2021-0137](https://doi.org/10.1515/reveh-2021-0137).
- [295] Kelly G. Lambert et al. “Brains in the city: Neurobiological effects of urbanization”. In: *Neuroscience & Biobehavioral Reviews* 58 (Nov. 2015), pp. 107–122. DOI: [10.1016/j.neubiorev.2015.04.007](https://doi.org/10.1016/j.neubiorev.2015.04.007).
- [296] C Neale et al. “The impact of walking in different urban environments on brain activity in older people”. In: *Cities & Health* 4.1 (Jan. 2020), pp. 94–106. DOI: [10.1080/23748834.2019.1619893](https://doi.org/10.1080/23748834.2019.1619893).
- [297] Bo Kyong Seo et al. “Homeownership, Depression, and Life Satisfaction in China: The Gender and Urban-Rural Disparities”. In: *International Journal of Environmental Research and Public Health* 19.22 (Nov. 2022), p. 14833. DOI: [10.3390/ijerph192214833](https://doi.org/10.3390/ijerph192214833).
- [298] Nan Zhang, James Nazroo, and Bram Vanhoutte. “The relationship between rural to urban migration in China and risk of depression in later life: An investigation of life course effects”. In: *Social Science & Medicine* 270 (Feb. 2021), p. 113637. DOI: [10.1016/j.socscimed.2020.113637](https://doi.org/10.1016/j.socscimed.2020.113637).
- [299] Chang Li and Bingchuan Sun. “Statistical analysis of mental influencing factors for anxiety and depression of rural and urban freshmen”. In: *Frontiers in Public Health* 11 (Dec. 2023). DOI: [10.3389/fpubh.2023.1235262](https://doi.org/10.3389/fpubh.2023.1235262).
- [300] Maximilian Nawrath et al. “Exploring uncharted territory: Do urban greenspaces support mental health in low- and middle-income countries?” In: *Environmental Research* 194 (Mar. 2021), p. 110625. DOI: [10.1016/j.envres.2020.110625](https://doi.org/10.1016/j.envres.2020.110625).

- [301] GN Bratman et al. “Nature experience reduces rumination and subgenual prefrontal cortex activation”. In: *Proceedings of the National Academy of Sciences* 112.28 (July 2015), pp. 8567–8572. DOI: [10.1073/pnas.1510459112](https://doi.org/10.1073/pnas.1510459112).
- [302] I Acosta et al. “Neuropsychiatric symptoms as risk factors of dementia in a Mexican population: A 10/66 Dementia Research Group study”. In: *Alzheimer’s & Dementia* 14.3 (Mar. 2018), pp. 271–279. DOI: [10.1016/j.jalz.2017.08.015](https://doi.org/10.1016/j.jalz.2017.08.015).
- [303] The World Bank. *The World by Income and Region*. 2023. URL: <https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html>.
- [304] Monica E. Nelson et al. “Cognitive Reserve, Alzheimer’s Neuropathology, and Risk of Dementia: A Systematic Review and Meta-Analysis”. In: *Neuropsychology Review* 31.2 (June 2021), pp. 233–250. DOI: [10.1007/s11065-021-09478-4](https://doi.org/10.1007/s11065-021-09478-4).
- [305] Hui Xu et al. “Association of lifespan cognitive reserve indicator with the risk of mild cognitive impairment and its progression to dementia”. In: *Alzheimer’s & Dementia* 16.6 (June 2020), pp. 873–882. DOI: [10.1002/alz.12085](https://doi.org/10.1002/alz.12085).
- [306] Wenzhe Yang et al. “Association of cognitive reserve with the risk of dementia in the UK Biobank: role of polygenic factors”. In: *The British Journal of Psychiatry* 224.6 (June 2024), pp. 213–220. DOI: [10.1192/bjp.2024.13](https://doi.org/10.1192/bjp.2024.13).
- [307] Jess Nithianantharajah and Anthony J. Hannan. “The neurobiology of brain and cognitive reserve: Mental and physical activity as modulators of brain disorders”. In: *Progress in Neurobiology* 89.4 (Dec. 2009), pp. 369–382. DOI: [10.1016/j.pneurobio.2009.10.001](https://doi.org/10.1016/j.pneurobio.2009.10.001).
- [308] Veronica Galvan and Dale Bredesen. “Neurogenesis in the Adult Brain: Implications for Alzheimers Disease”. In: *CNS & Neurological Disorders - Drug Targets* 6.5 (Oct. 2007), pp. 303–310. DOI: [10.2174/187152707783220938](https://doi.org/10.2174/187152707783220938).
- [309] S Dekhtyar et al. “A Life-Course Study of Cognitive Reserve in Dementia—From Childhood to Old Age”. In: *The American Journal of Geriatric Psychiatry* 23.9 (Sept. 2015), pp. 885–896. DOI: [10.1016/j.jagp.2015.02.002](https://doi.org/10.1016/j.jagp.2015.02.002).
- [310] SR de Rooij. “Are Brain and Cognitive Reserve Shaped by Early Life Circumstances?” In: *Frontiers in Neuroscience* 16 (June 2022). DOI: [10.3389/fnins.2022.825811](https://doi.org/10.3389/fnins.2022.825811).
- [311] EM Arenaza-Urquijo et al. “Sex and gender differences in cognitive resilience to aging and Alzheimer’s disease”. In: *Alzheimer’s & Dementia* 20 (July 2024), pp. 5695–5719. DOI: [10.1002/alz.13844](https://doi.org/10.1002/alz.13844).
- [312] Reuben N. Robbins et al. “Impact of urbanization on cognitive disorders”. In: *Current Opinion in Psychiatry* 32.3 (May 2019), pp. 210–217. DOI: [10.1097/YCO.0000000000000490](https://doi.org/10.1097/YCO.0000000000000490).
- [313] KR Gamble, JH Howard, and DV Howard. “Not Just Scenery: Viewing Nature Pictures Improves Executive Attention in Older Adults”. In: *Experimental Aging Research* 40.5 (Oct. 2014), pp. 513–530. DOI: [10.1080/0361073X.2014.956618](https://doi.org/10.1080/0361073X.2014.956618).
- [314] MG Berman, J Jonides, and S Kaplan. “The cognitive benefits of interacting with nature.” In: *Psychological science* 19.12 (Dec. 2008), pp. 1207–12. DOI: [10.1111/j.1467-9280.2008.02225.x](https://doi.org/10.1111/j.1467-9280.2008.02225.x).

- [315] J Ottosson and P Grahn. “A Comparison of Leisure Time Spent in a Garden with Leisure Time Spent Indoors: On Measures of Restoration in Residents in Geriatric Care”. In: *Landscape Research* 30.1 (Jan. 2005), pp. 23–55. DOI: [10.1080/0142639042000324758](https://doi.org/10.1080/0142639042000324758).
- [316] Heather Ohly et al. “Attention Restoration Theory: A systematic review of the attention restoration potential of exposure to natural environments”. In: *Journal of Toxicology and Environmental Health, Part B* 19.7 (Oct. 2016), pp. 305–343. DOI: [10.1080/10937404.2016.1196155](https://doi.org/10.1080/10937404.2016.1196155).
- [317] Kaplan S. “The Restorative Benefits of Nature - Toward an Integrative Framework.” In: *Journal of Environmental Psychology* 15 (1995), pp. 169–182.
- [318] Jenny Roe et al. “The Urban Built Environment, Walking and Mental Health Outcomes Among Older Adults: A Pilot Study”. In: *Frontiers in Public Health* 8 (Sept. 2020). DOI: [10.3389/fpubh.2020.575946](https://doi.org/10.3389/fpubh.2020.575946).
- [319] Stephen Kaplan and Marc G Berman. “Directed Attention as a Common Resource for Executive Functioning and Self-Regulation.” In: *Perspectives on psychological science : a journal of the Association for Psychological Science* 5.1 (Jan. 2010), pp. 43–57. DOI: [10.1177/1745691609356784](https://doi.org/10.1177/1745691609356784).
- [320] TR Herzog et al. “Reflection and Attentional Recovery as distinctive benefits of restorative environments”. In: *Journal of Environmental Psychology* 17.2 (June 1997), pp. 165–170. DOI: [10.1006/jevnp.1997.0051](https://doi.org/10.1006/jevnp.1997.0051).
- [321] P Ouellette, R Kaplan, and S Kaplan. “The monastery as a restorative environment”. In: *Journal of Environmental Psychology* 25.2 (June 2005), pp. 175–188. DOI: [10.1016/j.jenvp.2005.06.001](https://doi.org/10.1016/j.jenvp.2005.06.001).
- [322] Marica Cassarino, Isabella C. Tuohy, and Annalisa Setti. “Sometimes Nature Doesn’t Work: Absence of Attention Restoration in Older Adults Exposed to Environmental Scenes”. In: *Experimental Aging Research* 45.4 (Aug. 2019), pp. 372–385. DOI: [10.1080/0361073X.2019.1627497](https://doi.org/10.1080/0361073X.2019.1627497).
- [323] World Health Organisation. *Active ageing: A Policy Framework Active Ageing*. Tech. rep. Geneva: World Health Organisation (WHO), pp. 1–60. URL: <http://www.who.int/hpr/>.
- [324] J. K. Kim, M. Samaranayake, and S. Pradhan. “Epigenetic mechanisms in mammals”. In: *Cellular and Molecular Life Sciences* 66.4 (Feb. 2009), p. 596. DOI: [10.1007/s00018-008-8432-4](https://doi.org/10.1007/s00018-008-8432-4).
- [325] Sandro Galea, Monica Uddin, and Karestan Koenen. “The urban environment and mental disorders”. In: *Epigenetics* 6.4 (Apr. 2011), pp. 400–404. DOI: [10.4161/epi.6.4.14944](https://doi.org/10.4161/epi.6.4.14944).
- [326] RCE Bowyer et al. “Microbiota composition is moderately associated with greenspace composition in a UK cohort of twins”. In: *Science of The Total Environment* 813 (Mar. 2022), p. 152321. DOI: [10.1016/j.scitotenv.2021.152321](https://doi.org/10.1016/j.scitotenv.2021.152321).
- [327] Lilian Calderón-Garcidueñas et al. “Reduced repressive epigenetic marks, increased DNA damage and Alzheimer’s disease hallmarks in the brain of humans and mice exposed to particulate urban air pollution”. In: *Environmental Research* 183 (Apr. 2020), p. 109226. DOI: [10.1016/j.envres.2020.109226](https://doi.org/10.1016/j.envres.2020.109226).

- [328] R Peters et al. “Air Pollution and Dementia: A Systematic Review”. In: *Journal of Alzheimer’s Disease* 70.s1 (Aug. 2019), S145–S163. DOI: [10.3233/JAD-180631](https://doi.org/10.3233/JAD-180631).
- [329] Jonathan E. Thompson. “Airborne Particulate Matter”. In: *Journal of Occupational & Environmental Medicine* 60.5 (May 2018), pp. 392–423. DOI: [10.1097/JOM.0000000000001277](https://doi.org/10.1097/JOM.0000000000001277).
- [330] Hong Chen et al. “Living near major roads and the incidence of dementia, Parkinson’s disease, and multiple sclerosis: a population-based cohort study”. In: *The Lancet* 389.10070 (Feb. 2017), pp. 718–726. DOI: [10.1016/S0140-6736\(16\)32399-6](https://doi.org/10.1016/S0140-6736(16)32399-6).
- [331] Melinda C. Power et al. “Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: A systematic review of epidemiologic research”. In: *NeuroToxicology* 56 (Sept. 2016), pp. 235–253. DOI: [10.1016/j.neuro.2016.06.004](https://doi.org/10.1016/j.neuro.2016.06.004).
- [332] IM Carey et al. “Are noise and air pollution related to the incidence of dementia? A cohort study in London, England”. In: *BMJ Open* 8.9 (Sept. 2018), e022404. DOI: [10.1136/bmjopen-2018-022404](https://doi.org/10.1136/bmjopen-2018-022404).
- [333] Eirini Dimakakou et al. “Exposure to Environmental and Occupational Particulate Air Pollution as a Potential Contributor to Neurodegeneration and Diabetes: A Systematic Review of Epidemiological Research”. In: *International Journal of Environmental Research and Public Health* 15.8 (Aug. 2018), p. 1704. DOI: [10.3390/ijerph15081704](https://doi.org/10.3390/ijerph15081704).
- [334] B Bowe et al. “Burden of Cause-Specific Mortality Associated With PM 2.5 Air Pollution in the United States”. In: *JAMA Network Open* 2.11 (Nov. 2019), e1915834. DOI: [10.1001/jamanetworkopen.2019.15834](https://doi.org/10.1001/jamanetworkopen.2019.15834).
- [335] M Cacciottolo et al. “Particulate air pollutants, APOE alleles and their contributions to cognitive impairment in older women and to amyloidogenesis in experimental models”. In: *Translational Psychiatry* 7.1 (Jan. 2017), e1022–e1022. DOI: [10.1038/tp.2016.280](https://doi.org/10.1038/tp.2016.280).
- [336] Ai Milojevic et al. “Socioeconomic and urban-rural differentials in exposure to air pollution and mortality burden in England”. In: *Environmental Health* 16.1 (Dec. 2017), p. 104. DOI: [10.1186/s12940-017-0314-5](https://doi.org/10.1186/s12940-017-0314-5).
- [337] Adam Schuller and Luke Montrose. “Influence of Woodsmoke Exposure on Molecular Mechanisms Underlying Alzheimer’s Disease: Existing Literature and Gaps in Our Understanding”. In: *Epigenetics Insights* 13 (Jan. 2020), p. 251686572095487. DOI: [10.1177/2516865720954873](https://doi.org/10.1177/2516865720954873).
- [338] The Scottish Government. *Rural Scotland Key Facts 2021 People and Communities Services and Lifestyle Economy and Enterprise*. Tech. rep. Edinburgh: The Scottish Government, Feb. 2021. URL: <https://www.gov.scot/binaries/content/documents/govscot/publications/statistics/2021/02/rural-scotland-key-facts-2021/documents/rural-scotland-key-facts-2021/rural-scotland-key-facts-2021/govscot%3Adocument/rural-scotland-key-facts-2021.pdf>.
- [339] Adriana Elgueta. *BBC News: Leaving London: Pressure on outer boroughs as families move in*. Jan. 2024. URL: <https://www.bbc.co.uk/news/uk-england-london-67788531>.

- [340] Tom C. Russ et al. “Geographical variation in dementia: Systematic review with meta-analysis”. In: *International Journal of Epidemiology* 41.4 (Aug. 2012), pp. 1012–1032. DOI: [10.1093/ije/dys103](https://doi.org/10.1093/ije/dys103).
- [341] NIHR. *Prospero: International prospective register of systematic reviews*. URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=50323.
- [342] K Walesby et al. *NIHR PROSPERO International prospective register of systematic reviews: Geographical variation of dementia: systematic review with meta-analysis*. URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=50323.
- [343] EL Abner et al. “Rural-Urban Differences in Alzheimer’s Disease and Related Disorders Diagnostic Prevalence in Kentucky and West Virginia”. In: *The Journal of Rural Health* 32.3 (June 2016), pp. 314–320. DOI: [10.1111/jrh.12155](https://doi.org/10.1111/jrh.12155).
- [344] Momotazur Rahman et al. “Rural-urban differences in diagnostic incidence and prevalence of Alzheimer’s disease and related dementias”. In: *Alzheimer’s & Dementia* 17.7 (July 2021), pp. 1213–1230. DOI: [10.1002/alz.12285](https://doi.org/10.1002/alz.12285).
- [345] Lisa Ann Kirk Wiese et al. “Global rural health disparities in Alzheimer’s disease and related dementias: State of the science”. In: *Alzheimer’s & Dementia* 19.9 (Sept. 2023), pp. 4204–4225. DOI: [10.1002/alz.13104](https://doi.org/10.1002/alz.13104).
- [346] JR Roberts et al. “Dementia in rural settings: a scoping review exploring the personal experiences of people with dementia and their carers”. In: *Ageing and Society* (Mar. 2023), pp. 1–30. DOI: [10.1017/S0144686X2300003X](https://doi.org/10.1017/S0144686X2300003X).
- [347] D Morgan, A Innes, and J Kosteniuk. “Dementia care in rural and remote settings: A systematic review of formal or paid care”. In: *Maturitas* 68.1 (Jan. 2011), pp. 17–33. DOI: [10.1016/j.maturitas.2010.09.008](https://doi.org/10.1016/j.maturitas.2010.09.008).
- [348] Mandy Roheger et al. “Basic Diagnostic Work-Up Is More Complete in Rural than in Urban Areas for Patients with Dementia: Results of a Swedish Dementia Registry Study”. In: *Journal of Alzheimer’s Disease* 69.2 (May 2019), pp. 455–462. DOI: [10.3233/JAD-190017](https://doi.org/10.3233/JAD-190017).
- [349] Emma Nichols et al. “Differences in the measurement of cognition for the assessment of dementia across geographic contexts: Recommendations for cross-national research”. In: *Alzheimer’s & Dementia* 19.3 (Mar. 2023), pp. 1009–1019. DOI: [10.1002/alz.12740](https://doi.org/10.1002/alz.12740).
- [350] Tom C. Russ et al. “Geographical variation in dementia: Examining the role of environmental factors in Sweden and Scotland”. In: *Epidemiology* 26.2 (Mar. 2015), pp. 263–270. DOI: [10.1097/EDE.0000000000000230](https://doi.org/10.1097/EDE.0000000000000230).
- [351] Tom Russ et al. “Geographical Variation in Dementia Mortality in Italy, New Zealand, and Chile: The Impact of Latitude, Vitamin D, and Air Pollution”. In: *Dementia and Geriatric Cognitive Disorders* 42.1-2 (Sept. 2016), pp. 31–41. DOI: [10.1159/000447449](https://doi.org/10.1159/000447449).
- [352] N. Bagheri et al. “COMBINING GEOSPATIAL ANALYSIS WITH DEMENTIA RISK UTILISING GENERAL PRACTICE DATA: A SYSTEMATIC REVIEW”. In: *The Journal Of Prevention of Alzheimer’s Disease* (2017), pp. 1–7. DOI: [10.14283/jpad.2017.33](https://doi.org/10.14283/jpad.2017.33).

- [353] I Akushevich et al. “Geographic disparities in mortality from Alzheimer’s disease and related dementias”. In: *Journal of the American Geriatrics Society* 69.8 (Aug. 2021), pp. 2306–2315. DOI: [10.1111/jgs.17215](https://doi.org/10.1111/jgs.17215).
- [354] James Watson et al. “The Impact of Demographic, Socio-Economic and Geographic Factors on Mortality Risk among People Living with Dementia in England (2002–2016)”. In: *International Journal of Environmental Research and Public Health* 18.24 (Dec. 2021), p. 13405. DOI: [10.3390/ijerph182413405](https://doi.org/10.3390/ijerph182413405).
- [355] Peng Yin et al. “Temporal Trends and Geographic Variations in Dementia Mortality in China Between 2006 and 2012”. In: *Alzheimer Disease & Associated Disorders* 30.4 (Oct. 2016), pp. 348–353. DOI: [10.1097/WAD.0000000000000147](https://doi.org/10.1097/WAD.0000000000000147).
- [356] Alison Rataj et al. “Geographic Variation of Antidementia and Antipsychotic Medication Use Among US Nursing Home Residents With Dementia”. In: *Medical Care* (May 2024). DOI: [10.1097/MLR.0000000000002016](https://doi.org/10.1097/MLR.0000000000002016).
- [357] Aaron Reuben et al. “Dementia, dementia’s risk factors and premorbid brain structure are concentrated in disadvantaged areas: National register and birth-cohort geographic analyses”. In: *Alzheimer’s & Dementia* 20.5 (May 2024), pp. 3167–3178. DOI: [10.1002/alz.13727](https://doi.org/10.1002/alz.13727).
- [358] Neda Firouraghi et al. “The role of geographic information system and global positioning system in dementia care and research: a scoping review”. In: *International Journal of Health Geographics* 21.1 (Aug. 2022), p. 8. DOI: [10.1186/s12942-022-00308-1](https://doi.org/10.1186/s12942-022-00308-1).
- [359] Wendy Y. Xu et al. “Rural-Urban Disparities in Diagnosis of Early-Onset Dementia”. In: *JAMA Network Open* 5.8 (Aug. 2022), e2225805. DOI: [10.1001/jamanetworkopen.2022.25805](https://doi.org/10.1001/jamanetworkopen.2022.25805).
- [360] Ángela Gutiérrez et al. “Rural–Urban Health Care Cost Differences Among Latinx Adults With and Without Dementia in the United States”. In: *Journal of Aging and Health* (Oct. 2023). DOI: [10.1177/08982643231207517](https://doi.org/10.1177/08982643231207517).
- [361] Abhishek Menesgere et al. “Comparison of risk factors for dementia among rural and urban elderly adults-data from two cohort studies in India”. In: *Alzheimer’s & Dementia* 19.6 (June 2023), pp. 2443–2449. DOI: [10.1002/alz.12715](https://doi.org/10.1002/alz.12715).
- [362] Philip D St John et al. “Rural residence and risk of dementia.” In: *Canadian journal of rural medicine : the official journal of the Society of Rural Physicians of Canada = Journal canadien de la medecine rurale : le journal officiel de la Societe de medecine rurale du Canada* 21.3 (2016), pp. 73–9.
- [363] Momotazur Rahman et al. “Assessment of Rural-Urban Differences in Health Care Use and Survival Among Medicare Beneficiaries With Alzheimer Disease and Related Dementia”. In: *JAMA Network Open* 3.10 (Oct. 2020), e2022111. DOI: [10.1001/jamanetworkopen.2020.22111](https://doi.org/10.1001/jamanetworkopen.2020.22111).
- [364] Fei-fei Hu et al. “Population-attributable fractions of risk factors for all-cause dementia in China rural and urban areas: a cross-sectional study”. In: *Journal of Neurology* 269.6 (June 2022), pp. 3147–3158. DOI: [10.1007/s00415-021-10886-y](https://doi.org/10.1007/s00415-021-10886-y).
- [365] Thomas Astell-Burt and Xiaoqi Feng. “Is the risk of developing Alzheimer’s disease really higher in rural areas? A multilevel longitudinal study of 261,669 Australians aged 45 years and older tracked over 11 years”. In: *Health and Place* 54 (Nov. 2018), pp. 132–137. DOI: [10.1016/j.healthplace.2018.09.003](https://doi.org/10.1016/j.healthplace.2018.09.003).

- [366] Julie Kosteniuk et al. “Rural-urban differences in use of health services before and after dementia diagnosis: a retrospective cohort study”. In: *BMC Health Services Research* 24.1 (Mar. 2024), p. 399. DOI: [10.1186/s12913-024-10817-3](https://doi.org/10.1186/s12913-024-10817-3).
- [367] Meghan Zacher, Samantha Brady, and Susan E Short. “Geographic Patterns of Dementia in the United States: Variation by Place of Residence, Place of Birth, and Subpopulation”. In: *The Journals of Gerontology: Series B* 78.7 (June 2023), pp. 1192–1203. DOI: [10.1093/geronb/gbad045](https://doi.org/10.1093/geronb/gbad045).
- [368] M. Maria Glymour et al. “Geographic Distribution of Dementia Mortality”. In: *Alzheimer Disease & Associated Disorders* 25.3 (July 2011), pp. 196–202. DOI: [10.1097/WAD.0b013e31820905e7](https://doi.org/10.1097/WAD.0b013e31820905e7).
- [369] Kristen M George et al. “Association between birth in the stroke belt and late-life cognition with assessment of socioeconomic and cardiovascular risk factors: The Study of Healthy Aging in African Americans (STAR)”. In: *Alzheimer’s & Dementia* 17.S10 (Dec. 2021). DOI: [10.1002/alz.055715](https://doi.org/10.1002/alz.055715).
- [370] Paola Gilsanz et al. “Association Between Birth in a High Stroke Mortality State, Race, and Risk of Dementia”. In: *JAMA Neurology* 74.9 (Sept. 2017), p. 1056. DOI: [10.1001/jamaneurol.2017.1553](https://doi.org/10.1001/jamaneurol.2017.1553).
- [371] Melissa Lamar et al. “Relationship of Early-Life Residence and Educational Experience to Level and Change in Cognitive Functioning: Results of the Minority Aging Research Study”. In: *The Journals of Gerontology: Series B* 75.7 (Aug. 2020), e81–e92. DOI: [10.1093/geronb/gbz031](https://doi.org/10.1093/geronb/gbz031).
- [372] Michael Topping, Jinho Kim, and Jason Fletcher. “Geographic variation in Alzheimer’s disease mortality”. In: *PLOS ONE* 16.7 (July 2021), e0254174. DOI: [10.1371/journal.pone.0254174](https://doi.org/10.1371/journal.pone.0254174).
- [373] Michael Topping, Jinho Kim, and Jason Fletcher. “Association and pathways of birth in the stroke belt on old age dementia and stroke Mortality”. In: *SSM - Population Health* 15 (Sept. 2021), p. 100841. DOI: [10.1016/j.ssmph.2021.100841](https://doi.org/10.1016/j.ssmph.2021.100841).
- [374] Marica Cassarino and Annalisa Setti. “Complexity As Key to Designing Cognitive-Friendly Environments for Older People”. In: *Frontiers in Psychology* 7 (Aug. 2016). DOI: [10.3389/fpsyg.2016.01329](https://doi.org/10.3389/fpsyg.2016.01329).
- [375] GHSL - Global Human Settlement Layer. *The Degree of Urbanisation, a new global definition of cities, urban and rural areas*. URL: <https://human-settlement.emergency.copernicus.eu/degurba.php>.
- [376] Lewis Dijkstra et al. *Applying the Degree of Urbanisation – A methodological manual to define cities, towns and rural areas for international comparisons*. Tech. rep. 2021. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Applying_the_degree_of_urbanisation_manual.
- [377] Lewis Dijkstra and Hugo Poelman. *A harmonised definition of cities and rural areas: the new degree of urbanisation Working Papers*. Tech. rep.
- [378] Shlomo Angel et al. “The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050”. In: *Progress in Planning* 75.2 (Feb. 2011), pp. 53–107. DOI: [10.1016/j.progress.2011.04.001](https://doi.org/10.1016/j.progress.2011.04.001).

- [379] GeoPlace. *The Power of the Unique Property Reference Number (UPRN)*. 2024. URL: <https://www.geoplace.co.uk/addresses-streets/location-data/the-uprn>.
- [380] Yueheng Yin et al. “Epidemiology of Dementia in China in 2010–2020: A Systematic Review and Meta-Analysis”. In: *Healthcare* 12.3 (Jan. 2024), p. 334. DOI: [10.3390/healthcare12030334](https://doi.org/10.3390/healthcare12030334).
- [381] Fabiana Ribeiro et al. “Prevalence of dementia in Latin America and Caribbean countries: Systematic review and meta-analyses exploring age, sex, rurality, and education as possible determinants”. In: *Ageing Research Reviews* 81 (Nov. 2022), p. 101703. DOI: [10.1016/j.arr.2022.101703](https://doi.org/10.1016/j.arr.2022.101703).
- [382] C Dye. “Health and Urban Living”. In: *Science* 319.5864 (Feb. 2008), pp. 766–769. DOI: [10.1126/science.1150198](https://doi.org/10.1126/science.1150198).
- [383] O Gruebner et al. “Cities and Mental Health”. In: *Deutsches Ärzteblatt international* (Feb. 2017). DOI: [10.3238/arztebl.2017.0121](https://doi.org/10.3238/arztebl.2017.0121).
- [384] Naaheed Mukadam et al. “Changes in prevalence and incidence of dementia and risk factors for dementia: an analysis from cohort studies”. In: *The Lancet Public Health* 9.7 (July 2024), e443–e460. DOI: [10.1016/S2468-2667\(24\)00120-8](https://doi.org/10.1016/S2468-2667(24)00120-8).
- [385] World Health Organisation. *Global status report on the public health response to dementia*. 2021. ISBN: 978-92-4-003325-2. URL: <http://apps.who.int/>.
- [386] Jieun Lee et al. “Interventions to improve district-level routine health data in low-income and middle-income countries: a systematic review”. In: *BMJ Global Health* 6.6 (June 2021), e004223. DOI: [10.1136/bmjgh-2020-004223](https://doi.org/10.1136/bmjgh-2020-004223).
- [387] YW Hung et al. “Using routine health information data for research in low- and middle-income countries: a systematic review”. In: *BMC Health Services Research* 20.1 (Dec. 2020), p. 790. DOI: [10.1186/s12913-020-05660-1](https://doi.org/10.1186/s12913-020-05660-1).
- [388] Jim Todd and Michael Johnson Mahande. “Editorial: The Use of Routine Health Data in Low- and Middle-Income Countries.” In: *Frontiers in public health* 8 (2020), p. 413. DOI: [10.3389/fpubh.2020.00413](https://doi.org/10.3389/fpubh.2020.00413).
- [389] Benjamin Talbot et al. “Data Challenges in Addressing Chronic Kidney Disease in Low- and Lower-Middle-Income Countries”. In: *Kidney International Reports* 6.6 (June 2021), pp. 1503–1512. DOI: [10.1016/j.ekir.2021.03.901](https://doi.org/10.1016/j.ekir.2021.03.901).
- [390] Nena do Nascimento, Catherine Barker, and Isabel Brodsky. “Where is the evidence? The use of routinely-collected patient data to retain adults on antiretroviral treatment in low and middle income countries—a state of the evidence review”. In: *AIDS Care* 30.3 (Mar. 2018), pp. 267–277. DOI: [10.1080/09540121.2017.1381330](https://doi.org/10.1080/09540121.2017.1381330).
- [391] Yishu Liu et al. “Nation-Wide Routinely Collected Health Datasets in China: A Scoping Review”. In: *Public Health Reviews* 43 (Sept. 2022). DOI: [10.3389/phrs.2022.1605025](https://doi.org/10.3389/phrs.2022.1605025).
- [392] Zhu Chen. “Launch of the health-care reform plan in China”. In: *The Lancet* 373.9672 (Apr. 2009), pp. 1322–1324. DOI: [10.1016/S0140-6736\(09\)60753-4](https://doi.org/10.1016/S0140-6736(09)60753-4).
- [393] V. Jha, A. Y.-M. Wang, and H. Wang. “The impact of CKD identification in large countries: the burden of illness”. In: *Nephrology Dialysis Transplantation* 27.suppl 3 (Oct. 2012), pp. iii32–iii38. DOI: [10.1093/ndt/gfs113](https://doi.org/10.1093/ndt/gfs113).

- [394] SM Aljunid et al. “Health-Care Data Collecting, Sharing, and Using in Thailand, China Mainland, South Korea, Taiwan, Japan, and Malaysia”. In: *Value in Health* 15.1 (Jan. 2012), S132–S138. DOI: [10.1016/j.jval.2011.11.004](https://doi.org/10.1016/j.jval.2011.11.004).
- [395] KY Chan et al. “Tackling dementia globally: the Global Dementia Prevention Program (GloDePP) collaboration”. In: *Journal of Global Health* 9.2 (Dec. 2019). DOI: [10.7189/jogh.09.020103](https://doi.org/10.7189/jogh.09.020103).
- [396] SO Danso et al. “Application of Big Data and Artificial Intelligence technologies to dementia prevention research: an opportunity for low-and-middle-income countries”. In: *Journal of Global Health* 9.2 (Dec. 2019). DOI: [10.7189/jogh.09.020322](https://doi.org/10.7189/jogh.09.020322).
- [397] Shubash S Ganapathy et al. “Prevalence of dementia and quality of life of caregivers of people living with dementia in <scp>M</scp> alaysia”. In: *Geriatrics & Gerontology International* 20.S2 (Dec. 2020), pp. 16–20. DOI: [10.1111/ggi.14031](https://doi.org/10.1111/ggi.14031).
- [398] Blossom C M Stephan et al. “Prediction of dementia risk in low-income and middle-income countries (the 10/66 Study): an independent external validation of existing models”. In: *The Lancet Global Health* 8.4 (Apr. 2020), e524–e535. DOI: [10.1016/S2214-109X\(20\)30062-0](https://doi.org/10.1016/S2214-109X(20)30062-0).
- [399] G Yeo. “Association of Race and Ethnicity With Dementia”. In: *JAMA* 327.15 (Apr. 2022), p. 1454. DOI: [10.1001/jama.2022.4473](https://doi.org/10.1001/jama.2022.4473).
- [400] Shu Qin Li et al. “Dementia prevalence and incidence among the Indigenous and non-Indigenous populations of the Northern Territory”. In: *Medical Journal of Australia* 200.8 (May 2014), pp. 465–469. DOI: [10.5694/mja13.11052](https://doi.org/10.5694/mja13.11052).
- [401] P Byron et al. *Medicare and Pharmaceutical Benefits Scheme usage patterns in the Northern Territory 1993/94 to 2003/04*. Tech. rep. Department of Health and Community Services Northern Territory, 2005, p. 56. URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9fa338c9d8f583ece773974bb0412b783383532e>.
- [402] N Mukadam, C Cooper, and G Livingston. “Improving access to dementia services for people from minority ethnic groups”. In: *Current Opinion in Psychiatry* 26.4 (July 2013), pp. 409–414. DOI: [10.1097/YC0.0b013e32835ee668](https://doi.org/10.1097/YC0.0b013e32835ee668).
- [403] R Cheston et al. ““You just can’t do that in dementia care”: Barriers to partnership working within dementia services for people from South Asian communities”. In: *Dementia* (Sept. 2024). DOI: [10.1177/14713012241283189](https://doi.org/10.1177/14713012241283189).
- [404] N Mukadam et al. “Risk factors, ethnicity and dementia: A UK Biobank prospective cohort study of White, South Asian and Black participants”. In: *PLOS ONE* 17.10 (Oct. 2022), e0275309. DOI: [10.1371/journal.pone.0275309](https://doi.org/10.1371/journal.pone.0275309).
- [405] Alzheimer’s Society. *Black, Asian and minority ethnic communities and dementia research*. URL: <https://www.alzheimers.org.uk/for-researchers/black-asian-and-minority-ethnic-communities-and-dementia-research>.
- [406] S Griffiths et al. “Reflections on co-production: Developing a dementia research funding application with a diverse lived experience group”. In: *Dementia* 23.5 (July 2024), pp. 709–723. DOI: [10.1177/14713012241231916](https://doi.org/10.1177/14713012241231916).

- [407] J Reynolds, K Fryer, and C Mitchell. “Dementia and ethnicity — partnering with community organisations to co-produce research priorities and study design exploring access to dementia services for ethnic minority patients”. In: *British Journal of General Practice* 73.suppl 1 (July 2023), bjgp23X733701. DOI: [10.3399/bjgp23X733701](https://doi.org/10.3399/bjgp23X733701).
- [408] KM Mehta and GW Yeo. “Systematic review of dementia prevalence and incidence in United States race/ethnic populations”. In: *Alzheimer’s & Dementia* 13.1 (Jan. 2017), pp. 72–83. DOI: [10.1016/j.jalz.2016.06.2360](https://doi.org/10.1016/j.jalz.2016.06.2360).
- [409] Raj Kalara et al. “The 2022 symposium on dementia and brain aging in low- and middle-income countries: Highlights on research, diagnosis, care, and impact”. In: *Alzheimer’s & Dementia* 20.6 (June 2024), pp. 4290–4314. DOI: [10.1002/alz.13836](https://doi.org/10.1002/alz.13836).
- [410] InterTASC Information Specialists SubGroup (ISSG). *Geographic filters*. URL: <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home/geography>.
- [411] J Glanville et al. “So many filters, so little time: the development of a search filter appraisal checklist”. In: *Journal of the Medical Library Association : JMLA* 96.4 (Oct. 2008), pp. 356–361. DOI: [10.3163/1536-5050.96.4.011](https://doi.org/10.3163/1536-5050.96.4.011).
- [412] José M. Valderas et al. “Development of a Geographic Filter for PubMed to Identify Studies Performed in Spain”. In: *Revista Española de Cardiología (English Edition)* 59.12 (Jan. 2006), pp. 1244–1251. DOI: [10.1016/S1885-5857\(07\)60080-2](https://doi.org/10.1016/S1885-5857(07)60080-2).
- [413] Elizabeth Pienaar et al. “Developing a geographic search filter to identify randomised controlled trials in Africa: finding the optimal balance between sensitivity and precision”. In: *Health Information & Libraries Journal* 28.3 (Sept. 2011), pp. 210–215. DOI: [10.1111/j.1471-1842.2011.00936.x](https://doi.org/10.1111/j.1471-1842.2011.00936.x).
- [414] L Ayiku et al. “The MEDLINE UK filter: development and validation of a geographic search filter to retrieve research about the UK from OVID MEDLINE”. In: *Health Information & Libraries Journal* 34.3 (Sept. 2017), pp. 200–216. DOI: [10.1111/hir.12187](https://doi.org/10.1111/hir.12187).
- [415] L Ayiku et al. “How to Develop a Validated Geographic Search Filter: Five Key Steps”. In: *Evidence Based Library and Information Practice* 15.1 (Mar. 2020), pp. 170–178. DOI: [10.18438/ebliip29633](https://doi.org/10.18438/ebliip29633).
- [416] C Muentel et al. “Validated geographic search filters for bibliographic databases: a scoping review protocol”. In: *JBIM Evidence Synthesis* 22.3 (Mar. 2024), pp. 441–446. DOI: [10.11124/JBIES-23-00445](https://doi.org/10.11124/JBIES-23-00445).
- [417] L Ayiku et al. “The NICE UK geographic search filters for MEDLINE and Embase (Ovid): Post-development study to further evaluate precision and number-needed-to-read when retrieving UK evidence.” In: *Research synthesis methods* 11.5 (Sept. 2020), pp. 669–677. DOI: [10.1002/jrsm.1431](https://doi.org/10.1002/jrsm.1431).
- [418] L Ayiku, P Levay, and T Hudson. “The NICE OECD countries’ geographic search filters: Part 1—methodology for developing the draft MEDLINE and Embase (Ovid) filters”. In: *Journal of the Medical Library Association* 109.2 (July 2021). DOI: [10.5195/jmla.2021.978](https://doi.org/10.5195/jmla.2021.978).

- [419] L Ayiku et al. “The NICE OECD countries’ geographic search filters: Part 2—validation of the MEDLINE and Embase (Ovid) filters”. In: *Journal of the Medical Library Association* 109.4 (Nov. 2021). DOI: [10.5195/jmla.2021.1224](https://doi.org/10.5195/jmla.2021.1224).
- [420] A Sutton and F Campbell. “The SchHARR LMIC filter: Adapting a low- and middle-income countries geographic search filter to identify studies on preterm birth prevention and management”. In: *Research Synthesis Methods* 13.4 (July 2022), pp. 447–456. DOI: [10.1002/jrsm.1552](https://doi.org/10.1002/jrsm.1552).
- [421] Dawid Pieper et al. “Wie beschränken Autoren systematischer Reviews ihre Literatursuchen, wenn nur Studien aus Deutschland eingeschlossen werden sollen?” In: *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 117 (Nov. 2016), pp. 1–8. DOI: [10.1016/j.zefq.2016.08.007](https://doi.org/10.1016/j.zefq.2016.08.007).
- [422] J. P T Higgins. “Measuring inconsistency in meta-analyses”. In: *BMJ* 327.7414 (Sept. 2003), pp. 557–560. DOI: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557).
- [423] I Boutron et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Chapter 7: Considering bias and conflicts of interest among the included studies*. Ed. by JPT Higgins et al. Cochr, 2024. URL: www.training.cochrane.org/handbook..
- [424] JAC Sterne et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024): Chapter 25: Assessing risk of bias in a non-randomized study*. Ed. by JPT Higgins et al. Cochrane, 2024. URL: Available%20from%20www.training.cochrane.org/handbook..
- [425] Soo Young Kim et al. “Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity”. In: *Journal of Clinical Epidemiology* 66.4 (Apr. 2013), pp. 408–414. DOI: [10.1016/j.jclinepi.2012.09.016](https://doi.org/10.1016/j.jclinepi.2012.09.016).
- [426] JAC Sterne et al. “ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions”. In: *BMJ* (Oct. 2016), p. i4919. DOI: [10.1136/bmj.i4919](https://doi.org/10.1136/bmj.i4919).
- [427] GA Wells et al. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. Oxford, 2000.
- [428] Susan D. Shenkin et al. “Systematic reviews: guidance relevant for studies of older people”. In: *Age and Ageing* 46.5 (Sept. 2017), pp. 722–728. DOI: [10.1093/ageing/afx105](https://doi.org/10.1093/ageing/afx105).
- [429] JPT Higgins et al. “A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E)”. In: *Environment International* 186 (Apr. 2024), p. 108602. DOI: [10.1016/j.envint.2024.108602](https://doi.org/10.1016/j.envint.2024.108602).
- [430] Gordon H. Guyatt et al. “GRADE guidelines: 5. Rating the quality of evidence—publication bias”. In: *Journal of Clinical Epidemiology* 64.12 (Dec. 2011), pp. 1277–1282. DOI: [10.1016/j.jclinepi.2011.01.011](https://doi.org/10.1016/j.jclinepi.2011.01.011).
- [431] A Granholm, W Alhazzani, and MH Møller. “Use of the GRADE approach in systematic reviews and guidelines”. In: *British Journal of Anaesthesia* 123.5 (Nov. 2019), pp. 554–559. DOI: [10.1016/j.bja.2019.08.015](https://doi.org/10.1016/j.bja.2019.08.015).
- [432] Lifeng Lin and Haitao Chu. “Quantifying Publication Bias in Meta-Analysis”. In: *Biometrics* 74.3 (Sept. 2018), pp. 785–794. DOI: [10.1111/biom.12817](https://doi.org/10.1111/biom.12817).

- [433] EI Benchimol et al. “The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement”. In: *PLoS Medicine* 12.10 (2015). DOI: [10.1371/journal.pmed.1001885](https://doi.org/10.1371/journal.pmed.1001885).
- [434] Seblewengel Lemma et al. “Improving quality and use of routine health information system data in low- and middle-income countries: A scoping review”. In: *PLOS ONE* 15.10 (Oct. 2020), e0239683. DOI: [10.1371/journal.pone.0239683](https://doi.org/10.1371/journal.pone.0239683).
- [435] TJ Bollyky et al. “Lower-Income Countries That Face The Most Rapid Shift In Noncommunicable Disease Burden Are Also The Least Prepared”. In: *Health Affairs* 36.11 (Nov. 2017), pp. 1866–1875. DOI: [10.1377/hlthaff.2017.0708](https://doi.org/10.1377/hlthaff.2017.0708).
- [436] SIR Bhagyashree et al. “Diagnosis of Dementia by Machine learning methods in Epidemiological studies: a pilot exploratory study from south India”. In: *Social Psychiatry and Psychiatric Epidemiology* 53.1 (Jan. 2018), pp. 77–86. DOI: [10.1007/s00127-017-1410-0](https://doi.org/10.1007/s00127-017-1410-0).
- [437] Manish Kumar et al. “Research gaps in routine health information system design barriers to data quality and use in low- and middle-income countries: A literature review”. In: *The International Journal of Health Planning and Management* 33.1 (Jan. 2018). DOI: [10.1002/hpm.2447](https://doi.org/10.1002/hpm.2447).
- [438] Tim Wilkinson et al. “Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data”. In: *European Journal of Epidemiology* 34.6 (June 2019), pp. 557–565. DOI: [10.1007/s10654-019-00499-1](https://doi.org/10.1007/s10654-019-00499-1).
- [439] NIHR Maudsley Biomedical Research Centre. *Clinical Record Interactive Search (CRIS)*. 2024. URL: <https://www.maudsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/>.
- [440] NIHR Maudsley Biomedical Research Centre. *NIHR Maudsley Biomedical Research Centre: Clinical Record Interactive Search (CRIS) publications*. 2024. URL: <https://www.maudsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/cris-publications/>.
- [441] Natasha Chilman et al. “Understanding social and clinical associations with unemployment for people with schizophrenia and bipolar disorders: large-scale health records study”. In: *Social Psychiatry and Psychiatric Epidemiology* (Feb. 2024). DOI: [10.1007/s00127-024-02620-6](https://doi.org/10.1007/s00127-024-02620-6).
- [442] Axel A. S. Laurell et al. “Estimating demand for potential disease-modifying therapies for Alzheimer’s disease in the UK”. In: *The British Journal of Psychiatry* 224.6 (June 2024), pp. 198–204. DOI: [10.1192/bjp.2023.166](https://doi.org/10.1192/bjp.2023.166).
- [443] Chin-Kuo Chang et al. “Life Expectancy at Birth for People with Serious Mental Illness and Other Major Disorders from a Secondary Mental Health Care Case Register in London”. In: *PLoS ONE* 6.5 (May 2011), e19590. DOI: [10.1371/journal.pone.0019590](https://doi.org/10.1371/journal.pone.0019590).
- [444] Lisa A. Page et al. “Using electronic patient records to assess the impact of swine flu (influenza H1N1) on mental health patients”. In: *Journal of Mental Health* 20.1 (Feb. 2011), pp. 60–69. DOI: [10.3109/09638237.2010.542787](https://doi.org/10.3109/09638237.2010.542787).
- [445] James Rucker et al. “OPCRIT+: an electronic system for psychiatric diagnosis and data collection in clinical and research settings”. In: *British Journal of Psychiatry* 199.2 (Aug. 2011), pp. 151–155. DOI: [10.1192/bjp.bp.110.082925](https://doi.org/10.1192/bjp.bp.110.082925).

- [446] A Davison et al. *Public Attitudes to Data Integration Report prepared for Statistics New Zealand*. Tech. rep. New Zealand: Opus International Consultants Ltd., July 2015, pp. 1–80.
- [447] New Zealand Council for Civil Liberties. *New Zealand Council for Civil Liberties: Our Privacy Act rights and Stats Integrated Data Infrastructure*. June 2024. URL: <https://nzcccl.org.nz/our-privacy-act-rights-and-stats-integrated-data-infrastructure/>.
- [448] Health Data Research UK. *HDRUK: Involving and engaging patients and the public*. 2024. URL: <https://www.hdruk.ac.uk/about-us/involving-and-engaging-patients-and-the-public/>.
- [449] J Stockdale, J Cassell, and E Ford. ““Giving something back”: A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland”. In: *Wellcome Open Research* 3 (Jan. 2019), p. 6. DOI: [10.12688/wellcomeopenres.13531.2](https://doi.org/10.12688/wellcomeopenres.13531.2).
- [450] National Institute for Health and Care Research (NIHR). *Resources and training for public involvement in research*. 2024. URL: <https://www.learningforinvolvement.org.uk/>.
- [451] Aamnah Rahman et al. “Nothing about us, without us: is for us”. In: *Research Involvement and Engagement* 8.1 (Dec. 2022), p. 39. DOI: [10.1186/s40900-022-00372-8](https://doi.org/10.1186/s40900-022-00372-8).
- [452] Public Health Scotland. *NHS Scotland Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP)*. URL: <https://www.informationgovernance.scot.nhs.uk/pbpphsc/>.
- [453] NIHR et al. *UK Standards for Public Involvement*. 2024. URL: <https://sites.google.com/nihr.ac.uk/pi-standards/standards>.
- [454] J Beresford-Dent et al. “Embedding patient and public involvement in dementia research: Reflections from experiences during the ‘Journeying through Dementia’ randomised controlled trial”. In: *Dementia* 21.6 (Aug. 2022), pp. 1987–2003. DOI: [10.1177/14713012221106816](https://doi.org/10.1177/14713012221106816).
- [455] Scottish Dementia Working Research Sub-group UK. “Core principles for involving people with dementia in research: Innovative practice”. In: *Dementia* 13.5 (Sept. 2014), pp. 680–685. DOI: [10.1177/1471301214533255](https://doi.org/10.1177/1471301214533255).
- [456] J Miah et al. “Patient and public involvement in dementia research in the European Union: a scoping review”. In: *BMC Geriatrics* 19.1 (Dec. 2019), p. 220. DOI: [10.1186/s12877-019-1217-9](https://doi.org/10.1186/s12877-019-1217-9).
- [457] Dianne Gove et al. “Alzheimer Europe’s position on involving people with dementia in research through PPI (patient and public involvement)”. In: *Aging & Mental Health* 22.6 (June 2018), pp. 723–729. DOI: [10.1080/13607863.2017.1317334](https://doi.org/10.1080/13607863.2017.1317334).
- [458] A Burton, M Ogden, and C Cooper. “Planning and enabling meaningful patient and public involvement in dementia research”. In: *Current Opinion in Psychiatry* 32.6 (Nov. 2019), pp. 557–562. DOI: [10.1097/YCO.0000000000000548](https://doi.org/10.1097/YCO.0000000000000548).
- [459] Piotr Teodorowski et al. “Public Involvement and Engagement in Big Data Research: Scoping Review”. In: *Journal of Participatory Medicine* 16 (Aug. 2024), e56673. DOI: [10.2196/56673](https://doi.org/10.2196/56673).

- [460] Teodor Zidaru, Elizabeth M. Morrow, and Rich Stockley. “Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice”. In: *Health Expectations* 24.4 (Aug. 2021), pp. 1072–1124. DOI: [10.1111/hex.13299](https://doi.org/10.1111/hex.13299).
- [461] Crispin N. Goytia et al. “A Tale of 2 Constituencies”. In: *Medical Care* 56.Suppl 1 (Oct. 2018), S64–S69. DOI: [10.1097/MLR.0000000000000786](https://doi.org/10.1097/MLR.0000000000000786).
- [462] M Aitken et al. “Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies”. In: *BMC Medical Ethics* 17.1 (Dec. 2016), p. 73. DOI: [10.1186/s12910-016-0153-x](https://doi.org/10.1186/s12910-016-0153-x).
- [463] S Kalkman et al. “Patients’ and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence”. In: *Journal of Medical Ethics* 48.1 (Jan. 2022), pp. 3–13. DOI: [10.1136/medethics-2019-105651](https://doi.org/10.1136/medethics-2019-105651).
- [464] Nicola Howe et al. “Systematic review of participants’ attitudes towards data sharing: a thematic synthesis”. In: *Journal of Health Services Research & Policy* 23.2 (Apr. 2018), pp. 123–133. DOI: [10.1177/1355819617751555](https://doi.org/10.1177/1355819617751555).
- [465] Sigrid Sterckx et al. ““You hoped we would sleep walk into accepting the collection of our data”: controversies surrounding the UK care.data scheme and their wider relevance for biomedical research”. In: *Medicine, Health Care and Philosophy* 19.2 (June 2016), pp. 177–190. DOI: [10.1007/s11019-015-9661-6](https://doi.org/10.1007/s11019-015-9661-6).
- [466] Tracy Jackson et al. “Patient and public involvement in research: from tokenistic box ticking to valued team members”. In: *BMC Medicine* 18.1 (Dec. 2020), p. 79. DOI: [10.1186/s12916-020-01544-7](https://doi.org/10.1186/s12916-020-01544-7).
- [467] S Staniszewska et al. “GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research”. In: *BMJ* (Aug. 2017), j3453. DOI: [10.1136/bmj.j3453](https://doi.org/10.1136/bmj.j3453).
- [468] World Health Organisation. *World Health Organisation Constitution*. URL: <https://www.who.int/about/governance/constitution>.
- [469] Måns Rosén. “National Health Data Registers: a Nordic heritage to public health”. In: *Scandinavian Journal of Public Health* 30.2 (June 2002), pp. 81–85. DOI: [10.1080/140349401753683444](https://doi.org/10.1080/140349401753683444).
- [470] John Maret-Ouda et al. “Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data”. In: *Scandinavian Journal of Public Health* 45.17_suppl (July 2017), pp. 14–19. DOI: [10.1177/1403494817702336](https://doi.org/10.1177/1403494817702336).
- [471] NHS England. *NHS England: SNOMED CT*. URL: <https://www.england.nhs.uk/digitaltechnology/digital-primary-care/snomed-ct/>.
- [472] NHS. *NHS British National Formulary (BNF) Chapter: Code search*. URL: <https://dmd-browser.nhsbsa.nhs.uk/code-search-bnf>.
- [473] NHS Business Services Authority. *NHS Business Services Authority: BNF SNOMED Mapping*. URL: <https://www.nhsbsa.nhs.uk/prescription-data/understanding-our-data/bnf-snomed-mapping>.
- [474] Health New Zealand Te Whatu Ora. *Introduction to Classification*. Aug. 2024. URL: <https://www.tewhatuora.govt.nz/for-health-professionals/data-and-statistics/classification-information/introduction>.

- [475] SNOMED International. *SNOMED International*. URL: <https://www.snomed.org/>.
- [476] Health Informatics New Zealand. *World Health Assembly - New Zealand joins statement challenging WHO report on ICD-11*. May 2023. URL: <https://www.hinz.org.nz/news/641562/World-Health-Assembly---New-Zealand-joins-statement-challenging-WHO-report-on-ICD-11.htm>.
- [477] Mohammad Chowdhury et al. “Use of Machine Learning and Artificial Intelligence Methods in Geriatric Mental Health Research Involving Electronic Health Record or Administrative Claims Data: A Systematic Review”. In: *Frontiers in Psychiatry* 12 (Sept. 2021). DOI: [10.3389/fpsy.2021.738466](https://doi.org/10.3389/fpsy.2021.738466).
- [478] Katie Harron, Harvey Goldstein, and Chris Dibben. *Methodological Developments in Data Linkage*. 1st. London: Wiley, Dec. 2015, pp. 1–288.
- [479] HL Dunn. “Record Linkage.” In: *American journal of public health and the nation’s health* 36.12 (Dec. 1946), pp. 1412–6.
- [480] Morten Schmidt, Lars Pedersen, and Henrik Toft Sørensen. “The Danish Civil Registration System as a tool in epidemiology”. In: *European Journal of Epidemiology* 29.8 (Aug. 2014), pp. 541–549. DOI: [10.1007/s10654-014-9930-3](https://doi.org/10.1007/s10654-014-9930-3).
- [481] Sigrun Alba Johannesdottir et al. “Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions”. In: *Clinical Epidemiology* (Nov. 2012), p. 303. DOI: [10.2147/CLEP.S37587](https://doi.org/10.2147/CLEP.S37587).
- [482] Public Health Scotland. *CHI Linkage and Indexing (CHILI): Community Health Index (CHI)*. URL: <https://publichealthscotland.scot/services/chi-linkage-and-indexing-chili/about-the-chi-linkage-and-indexing-team-chili/community-health-index-chi/>.
- [483] Public Health Scotland. *CHI Linkage and Indexing (CHILI): Services*. URL: <https://publichealthscotland.scot/services/chi-linkage-and-indexing-chili/services/>.
- [484] National Institute for Health and Care Research (NIHR). *Clinical Practice Research Datalink*. URL: <https://www.cprd.com/>.
- [485] NHS Digital. *Data technology and improves lives*. URL: <https://digital.nhs.uk/>.
- [486] NHS England. *NHS England Secure Data Environment*. URL: <https://digital.nhs.uk/services/secure-data-environment-service>.
- [487] NHS England Digital. *Spine*. URL: <https://digital.nhs.uk/services/spine#what-information-is-available>.
- [488] University of Oxford Bennett Institute for Applied Data Science. *OpenSAFELY*. 2025. URL: <https://www.opensafely.org/>.
- [489] Ronan A Lyons et al. “The SAIL databank: linking multiple health and social care datasets”. In: *BMC Medical Informatics and Decision Making* 9.1 (Dec. 2009), p. 3. DOI: [10.1186/1472-6947-9-3](https://doi.org/10.1186/1472-6947-9-3).
- [490] SAIL Databank. *SAIL Databank*. URL: <https://saildatabank.com/>.
- [491] HSC Business Services Organisation. *Honest Broker Service*. URL: <https://bso.hscni.net/directorates/digital/honest-broker-service/>.

- [492] Rasmus Gregersen, Rikke Wiingreen, and Jacob Rosenberg. “[Health-related register-based research in Denmark].” In: *Ugeskrift for læger* 180.43 (Oct. 2018).
- [493] Margaret K. Saunders. “In Denmark, Big Data Goes To Work”. In: *Health Affairs* 33.7 (July 2014), pp. 1245–1245. DOI: [10.1377/hlthaff.2014.0513](https://doi.org/10.1377/hlthaff.2014.0513).
- [494] Barry J Milne et al. “Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI)”. In: *International Journal of Epidemiology* 48.3 (June 2019), pp. 677–677. DOI: [10.1093/ije/dyz014](https://doi.org/10.1093/ije/dyz014).
- [495] Rosemary Wyber et al. “Mégadonnées dans le domaine de la santé mondiale: Amélioration de la santé dans les pays à revenu faible et intermédiaire”. In: *Bulletin of the World Health Organization* 93.3 (2015), pp. 203–208. DOI: [10.2471/BLT.14.139022](https://doi.org/10.2471/BLT.14.139022).
- [496] Kari Furu et al. “The Nordic Countries as a Cohort for Pharmacoepidemiological Research”. In: *Basic & Clinical Pharmacology & Toxicology* 106.2 (Feb. 2010), pp. 86–94. DOI: [10.1111/j.1742-7843.2009.00494.x](https://doi.org/10.1111/j.1742-7843.2009.00494.x).
- [497] JK Burton et al. “Identifying care-home residents in routine healthcare datasets: a diagnostic test accuracy study of five methods”. In: *Age and Ageing* 48.1 (Jan. 2019), pp. 114–121. DOI: [10.1093/ageing/afy137](https://doi.org/10.1093/ageing/afy137).
- [498] Fatemeh Torabi et al. “Common governance model: a way to avoid data segregation between existing trusted research environment”. In: *International Journal of Population Data Science* 8.4 (Nov. 2023). DOI: [10.23889/ijpds.v8i4.2164](https://doi.org/10.23889/ijpds.v8i4.2164).
- [499] Becky Field et al. “Recruiting hard to reach populations to studies: breaking the silence: an example from a study that recruited people with dementia”. In: *BMJ Open* 9.11 (Nov. 2019), e030829. DOI: [10.1136/bmjopen-2019-030829](https://doi.org/10.1136/bmjopen-2019-030829).
- [500] Joan K Morris et al. “EUROlinkCAT protocol for a European population-based data linkage study investigating the survival, morbidity and education of children with congenital anomalies”. In: *BMJ Open* 11.6 (June 2021), e047859. DOI: [10.1136/bmjopen-2020-047859](https://doi.org/10.1136/bmjopen-2020-047859).
- [501] Le Gao et al. “Linking cohort-based data with electronic health records: a proof-of-concept methodological study in Hong Kong”. In: *BMJ Open* 11.6 (June 2021), e045868. DOI: [10.1136/bmjopen-2020-045868](https://doi.org/10.1136/bmjopen-2020-045868).
- [502] IJ Deary et al. “Cohort Profile: The Lothian Birth Cohorts of 1921 and 1936”. In: *International Journal of Epidemiology* 41.6 (Dec. 2012), pp. 1576–1584. DOI: [10.1093/ije/dyr197](https://doi.org/10.1093/ije/dyr197).
- [503] S Bauermeister et al. “The Dementias Platform UK (DPUK) Data Portal”. In: *European Journal of Epidemiology* 35.6 (June 2020), pp. 601–611. DOI: [10.1007/s10654-020-00633-4](https://doi.org/10.1007/s10654-020-00633-4).
- [504] Hannes Ulrich et al. “Understanding the Nature of Metadata: Systematic Review”. In: *Journal of Medical Internet Research* 24.1 (Jan. 2022), e25440. DOI: [10.2196/25440](https://doi.org/10.2196/25440).
- [505] MD Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3.1 (Mar. 2016), p. 160018. DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- [506] Angela Wood et al. “Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource”. In: *BMJ* (Apr. 2021), n826. DOI: [10.1136/bmj.n826](https://doi.org/10.1136/bmj.n826).

- [507] World Health Organisation. *Geolocated Health Facilities Data Initiative*. URL: <https://www.who.int/data/GIS/GHFD>.
- [508] UK Statistics Authority and Office for Statistics Regulation. *Code of Practice for Statistics: Data governance*. 2024. URL: <https://code.statisticsauthority.gov.uk/the-code/trustworthiness/t6-data-governance/>.
- [509] DataGrail. *What are the 4 data governance pillars?* July 2023. URL: <https://www.datagrail.io/blog/data-privacy/data-governance-pillars/>.
- [510] Information Commissioner’s Office (ICO). *UK GDPR guidance and resources*. URL: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/>.
- [511] UK Legislation. *Data Protection Act 2018*. URL: <https://www.legislation.gov.uk/ukpga/2018/12/contents>.
- [512] UK Data Service. *What is the Five Safes Framework?* 2024. URL: <https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/>.
- [513] UK Legislation. *National Health Service Act 2006 Section 251*. Mar. 2007. URL: <https://www.legislation.gov.uk/ukpga/2006/41/section/251>.
- [514] NHS Health Research Authority. *Integrated Research Application System (IRAS)*. URL: <https://www.hra.nhs.uk/about-us/committees-and-services/integrated-research-application-system/>.
- [515] E Von Elm et al. “The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies”. In: *Ann Intern Med* 147 (2007), pp. 573–577. URL: www.annals.org.
- [516] Sinéad M Langan et al. “The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE)”. In: *BMJ* (Nov. 2018), k3532. DOI: [10.1136/bmj.k3532](https://doi.org/10.1136/bmj.k3532).
- [517] Mahrukh Imran et al. “Methods and results used in the development of a consensus-driven extension to the Consolidated Standards of Reporting Trials (CONSORT) statement for trials conducted using cohorts and routinely collected data (CONSORT-ROUTINE)”. In: *BMJ Open* 11.4 (Apr. 2021), e049093. DOI: [10.1136/bmjopen-2021-049093](https://doi.org/10.1136/bmjopen-2021-049093).
- [518] Ruth A. Sibbett et al. “Dementia ascertainment using existing data in UK longitudinal and cohort studies: A systematic review of methodology”. In: *BMC Psychiatry* 17.1 (July 2017). DOI: [10.1186/s12888-017-1401-4](https://doi.org/10.1186/s12888-017-1401-4).
- [519] S Horrocks et al. “Accuracy of routinely-collected healthcare data for identifying motor neurone disease cases: A systematic review”. In: *PLOS ONE* 12.2 (Feb. 2017), e0172639. DOI: [10.1371/journal.pone.0172639](https://doi.org/10.1371/journal.pone.0172639).
- [520] R. Liisa Jaakkimainen et al. “Identification of physician-diagnosed Alzheimer’s disease and related dementias in population-based administrative data: A validation study using family physicians’ electronic medical records”. In: *Journal of Alzheimer’s Disease* 54.1 (Aug. 2016), pp. 337–349. DOI: [10.3233/JAD-160105](https://doi.org/10.3233/JAD-160105).

- [521] National Records of Scotland. *The Medical Certificate of the Cause of Death (MCCD)*. URL: <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths/deaths-background-information/death-certificates-and-coding-the-causes-of-death/the-medical-certificate-of-the-cause-of-death>.
- [522] UK Government: Department of Health and Social Care. *Medical certificate of cause of death (MCCD): guidance for medical practitioners*. URL: <https://www.gov.uk/government/publications/medical-certificate-of-cause-of-death-mccd-guidance-for-medical-practitioners>.
- [523] Lu Gao et al. “Accuracy of death certification of dementia in population-based samples of older people: Analysis over time”. In: *Age and Ageing* 47.4 (July 2018), pp. 589–594. DOI: [10.1093/ageing/afy068](https://doi.org/10.1093/ageing/afy068).
- [524] NHS. *Death certification processes: information for medical practitioners after the Coronavirus Act 2020 expires*. Tech. rep. 2022.
- [525] Andrew Sommerlad et al. “Accuracy of general hospital dementia diagnoses in England: Sensitivity, specificity, and predictors of diagnostic accuracy 2008–2016”. In: *Alzheimer’s and Dementia* 14.7 (July 2018), pp. 933–943. DOI: [10.1016/j.jalz.2018.02.012](https://doi.org/10.1016/j.jalz.2018.02.012).
- [526] H Wetterberg et al. “The Effect of Diagnostic Criteria on Dementia Prevalence – A Population-Based Study From Gothenburg, Sweden”. In: *The American Journal of Geriatric Psychiatry* 32.2 (Feb. 2024), pp. 230–243. DOI: [10.1016/j.jagp.2023.08.018](https://doi.org/10.1016/j.jagp.2023.08.018).
- [527] LA McGuinness et al. “The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: A systematic review”. In: *Pharmacoepidemiology and Drug Safety* 28.2 (Feb. 2019), pp. 244–255. DOI: [10.1002/pds.4669](https://doi.org/10.1002/pds.4669).
- [528] A Brown et al. “Comparison of dementia recorded in routinely collected hospital admission data in England with dementia recorded in primary care”. In: *Emerging Themes in Epidemiology* 13.1 (Oct. 2016), pp. 1–9. DOI: [10.1186/s12982-016-0053-z](https://doi.org/10.1186/s12982-016-0053-z).
- [529] Mar Pujades-Rodriguez et al. “The diagnosis, burden and prognosis of dementia: A record-linkage cohort study in England”. In: *PLoS ONE* 13.6 (June 2018). DOI: [10.1371/journal.pone.0199026](https://doi.org/10.1371/journal.pone.0199026).
- [530] Donncha S. Mullin et al. “Identifying dementia using medical data linkage in a longitudinal cohort study: Lothian Birth Cohort 1936”. In: *BMC Psychiatry* 23.1 (May 2023), p. 303. DOI: [10.1186/s12888-023-04797-7](https://doi.org/10.1186/s12888-023-04797-7).
- [531] National Institute for Health and Care Excellence (NICE). *Dementia*. URL: <https://bnf.nice.org.uk/treatment-summaries/dementia/>.
- [532] Daniel John Exeter, Sarah Rodgers, and Clive Eric Sabel. ““Whose data is it anyway?” The implications of putting small area-level health and social data online”. In: *Health Policy* 114.1 (Jan. 2014), pp. 88–96. DOI: [10.1016/j.healthpol.2013.07.012](https://doi.org/10.1016/j.healthpol.2013.07.012).
- [533] Geospatial Commission. *The UK’s Geospatial Strategy, 2020 to 2025: Unlocking the power of location*. Tech. rep. 2020, pp. 1–71. URL: https://assets.publishing.service.gov.uk/media/5ef322cae90e075c52d059f1/Geospatial_Strategy.pdf.

- [534] UK Government: Geospatial Commission. *Unlocking the power of location: The UK's geospatial strategy 2020 to 2025*. 2020. URL: <https://www.gov.uk/government/publications/unlocking-the-power-of-locationthe-uks-geospatial-strategy/b0ff7327-0c83-4a26-81ad-3b66ae004a5d>.
- [535] Gillian Harper et al. "Determining households from patient addresses and unique property reference numbers in general practitioner electronic health records". In: *International Journal of Population Data Science* 9.1 (Apr. 2024). DOI: [10.23889/ijpds.v9i1.2379](https://doi.org/10.23889/ijpds.v9i1.2379).
- [536] DataLoch. *DataLoch*. URL: <https://dataloch.org/data>.
- [537] C Sudlow. *The Sudlow Review: Unifying Health Data in the UK*. URL: <https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/>.
- [538] Health Data Research UK. *Health Data Research UK (HDR UK)*. URL: <https://www.hdruk.ac.uk/>.
- [539] Health Data Research UK. "Health-Data-Research-Hub-Network". In: (). URL: <https://www.hdruk.ac.uk/wp-content/uploads/2022/08/Health-Data-Research-Hub-Network.pdf>.
- [540] DARE UK. *What is DARE UK?* URL: <https://dareuk.org.uk/about/>.
- [541] Digital Health and Care Scotland. *New Community Health Index (CHI) system*. URL: <https://www.digihealthcare.scot/delivery-plan-2024-25/new-chi/>.
- [542] Public Health Scotland (PHS). *Electronic Data Research and Innovation Service (eDRIS)*. Nov. 2023. URL: <https://publichealthscotland.scot/services/data-research-and-innovation-services/electronic-data-research-and-innovation-service-edris/overview/what-is-edris/>.
- [543] Research Data Scotland. *Research Data Scotland: About RDS*. 2024. URL: <https://www.researchdata.scot/about-rds/>.
- [544] Scottish Government. *Scottish Government Statistics: request our data*. URL: <https://www.gov.scot/publications/scottish-government-statistics-request-our-data/>.
- [545] Research Data Scotland. *PBPP Approval Pathway*. 2024. URL: <https://www.researchdata.scot/accessing-data/researcher-access-service/pbpp-approval-pathway/>.
- [546] The Scottish Government. *Enabling, Connecting and Empowering: Care in the Digital Age*. Tech. rep. The Scottish Government, 2021, pp. 1–33. URL: <https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2021/10/scotlands-digital-health-care-strategy/documents/enabling-connecting-empowering-care-digital-age/enabling-connecting-empowering-care-digital-age/govscot%3Adocument/enabling-connecting-empowering-care-digital-age.pdf>.

- [547] The Scottish Government. *Greater access, better insight, improved outcomes: a strategy for data-driven care in the digital age* Scottish Government and COSLA. Tech. rep. URL: <https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2023/02/data-strategy-health-social-care-2/documents/greater-access-better-insight-improved-outcomes-strategy-data-driven-care-digital-age/greater-access-better-insight-improved-outcomes-strategy-data-driven-care-digital-age/govscot%3Adocument/greater-access-better-insight-improved-outcomes-strategy-data-driven-care-digital-age.pdf>.
- [548] NHS Scotland. "The SNOMED CT Implementation Programme is leading NHS Scotland's healthcare terminology transformation". In: (). URL: <https://www.snomed.nhs.scot/>.
- [549] The Scottish Government. *Care in the Digital Age: Delivery Plan 2022-23*. Tech. rep. The Scottish Government, Nov. 2022. URL: <https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2022/11/care-digital-age-delivery-plan2/documents/care-digital-age-delivery-plan-2022-23/care-digital-age-delivery-plan-2022-23/govscot%3Adocument/care-digital-age-delivery-plan-2022-23.pdf>.
- [550] The Scottish Government and COSLA. *Care in the Digital Age: Delivery Plan 2024-25*. Tech. rep. The Scottish Government, Apr. 2024, pp. 1–36. URL: <https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2024/05/care-digital-age-delivery-plan-2024-25/documents/care-digital-age-delivery-plan-2024-25-scotlands-digital-health-care-strategy/care-digital-age-delivery-plan-2024-25-scotlands-digital-health-care-strategy/govscot%3Adocument/care-digital-age-delivery-plan-2024-25-scotlands-digital-health-care-strategy.pdf>.
- [551] University of Dundee. *Health Informatics Centre (HIC)*. URL: <https://www.dundee.ac.uk/hic>.
- [552] DataLoch. *Unveiling the Impact of the SARA Project: collaboratively improving health data access*. URL: <https://dataloch.org/insights/news/unveiling-impact-sara-project-collaboratively-improving-health-data-access>.
- [553] Research Data Scotland. *Researcher Access Service*. 2024. URL: <https://www.researchdata.scot/accessing-data/researcher-access-service/>.
- [554] Adrian Hopper. *Geriatric Medicine GIRFT Programme National Specialty Report February 2021 GIRFT is delivered in partnership with the Royal National Orthopaedic Hospital NHS Trust, NHS England and NHS Improvement 2*. Tech. rep.
- [555] EPIC. *EPIC*. URL: <https://www.epic.com/about/>.
- [556] Epic. *MyChart is Epic*. URL: <https://www.mychart.org/>.
- [557] Australian Digital Health Authority. *My Health Record*. 2024. URL: <https://www.digitalhealth.gov.au/initiatives-and-programs/my-health-record>.
- [558] SCI-Diabetes NHS Scotland. *Scottish Care Information Diabetes Collaboration (SCI-Diabetes)*. 2002. URL: <https://www.sci-diabetes.scot.nhs.uk/>.

- [559] British Geriatrics Society (BGS) Ageing Data Research Collaborative. *Ageing Data Research Collaborative (ADRC)*. 2020. URL: <https://www.bgs.org.uk/resources/ageing-data-research-collaborative-geridata>.
- [560] European Geriatric Medicine Society (EuGMS) Big Data Special Interest Group (SIG). *European Geriatric Medicine Society: Big Data*. URL: <https://www.eugms.org/research-cooperation/special-interest-groups/big-data.html>.
- [561] Professional Record Standards Body. *Professional Record Standards Body*. URL: <https://theprsb.org/>.
- [562] Jennifer Kirsty Burton et al. “Developing a minimum data set for older adult care homes in the UK: exploring the concept and defining early core principles”. In: *The Lancet Healthy Longevity* 3.3 (Mar. 2022), e186–e193. DOI: [10.1016/S2666-7568\(22\)00010-1](https://doi.org/10.1016/S2666-7568(22)00010-1).
- [563] Massirfulay Kpehe Musa et al. “The uptake and use of a minimum data set (MDS) for older people living and dying in care homes: a realist review”. In: *BMC Geriatrics* 22.1 (Jan. 2022), p. 33. DOI: [10.1186/s12877-021-02705-w](https://doi.org/10.1186/s12877-021-02705-w).
- [564] Ann-Marie Towers et al. “Piloting of a minimum data set for older people living in care homes in England: protocol for a longitudinal, mixed-methods study”. In: *BMJ Open* 13.2 (Feb. 2023), e071686. DOI: [10.1136/bmjopen-2023-071686](https://doi.org/10.1136/bmjopen-2023-071686).
- [565] Barbara Hanratty et al. “Covid-19 and lack of linked datasets for care homes”. In: *BMJ* (June 2020), p. m2463. DOI: [10.1136/bmj.m2463](https://doi.org/10.1136/bmj.m2463).
- [566] Statistics New Zealand Tatauranga Aotearoa. *National population estimates: At 30 June 2023*. June 2023. URL: <https://www.stats.govt.nz/information-releases/national-population-estimates-at-30-june-2023>.
- [567] Scottish Government. *New dementia strategy for Scotland: Everyone’s Story - Building blocks*. 2023. URL: <https://www.gov.scot/publications/new-dementia-strategy-scotland-everyones-story/pages/7/>.
- [568] E Ma’u et al. *Dementia Economic Impact Report 2020*. Tech. rep. Auckland, New Zealand: University of Auckland, 2021. URL: <https://cdn.alzheimers.org.nz/wp-content/uploads/2021/09/Dementia-Economic-Impact-Report-2020.pdf>.
- [569] Data Saves Lives. *Data Saves Lives*. URL: <https://datasaveslives.eu/>.
- [570] Esther S. Yao et al. “Demographic discrepancies between administrative-prioritisation and self-prioritisation of multiple ethnic identifications”. In: *Social Science Research* 103 (Mar. 2022), p. 102648. DOI: [10.1016/j.ssresearch.2021.102648](https://doi.org/10.1016/j.ssresearch.2021.102648).
- [571] A Morning. “Ethnic Classification in Global Perspective: A Cross-National Survey of the 2000 Census Round”. In: *Population Research and Policy Review* 27.2 (Apr. 2008), pp. 239–272. DOI: [10.1007/s11113-007-9062-5](https://doi.org/10.1007/s11113-007-9062-5).
- [572] A Smedley and BD Smedley. “Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race.” In: *American Psychologist* 60.1 (2005), pp. 16–26. DOI: [10.1037/0003-066X.60.1.16](https://doi.org/10.1037/0003-066X.60.1.16).
- [573] ES Yao et al. “Effects of Ethnic Classification on Substantive Findings in Adolescent Mental Health Outcomes”. In: *Journal of Youth and Adolescence* 51.8 (Aug. 2022), pp. 1581–1596. DOI: [10.1007/s10964-022-01612-6](https://doi.org/10.1007/s10964-022-01612-6).

- [574] WD Roth. “The multiple dimensions of race”. In: *Ethnic and Racial Studies* 39.8 (June 2016), pp. 1310–1338. DOI: [10.1080/01419870.2016.1140793](https://doi.org/10.1080/01419870.2016.1140793).
- [575] WD Roth. “The End of the One-Drop Rule? Labeling of Multiracial Children in Black Intermarriages”. In: *Sociological Forum* 20.1 (Mar. 2005), pp. 35–67. DOI: [10.1007/s11206-005-1897-0](https://doi.org/10.1007/s11206-005-1897-0).
- [576] Nikki Khanna. “Multiracial Americans: Racial Identity Choices and Implications for the Collection of Race Data”. In: *Sociology Compass* 6.4 (Apr. 2012), pp. 316–331. DOI: [10.1111/j.1751-9020.2011.00454.x](https://doi.org/10.1111/j.1751-9020.2011.00454.x).
- [577] P Callister et al. “Measuring Ethnicity in New Zealand: Developing Tools for Health Outcomes Analysis”. In: *Ethnicity & Health* 12.4 (Sept. 2007), pp. 299–320. DOI: [10.1080/13557850701300699](https://doi.org/10.1080/13557850701300699).
- [578] PA Carr et al. “Ethnic identification complexity across generations: Evidence from growing up in new zealand: Official newsletter of the new zealand demographic society”. In: *New Zealand Population Review* 43 (2017), pp. 35–61. URL: <https://www.proquest.com/scholarly-journals/ethnic-identification-complexity-across/docview/2052768485/se-2>.
- [579] VM Mays et al. “Classification of Race and Ethnicity: Implications for Public Health”. In: *Annual Review of Public Health* 24.1 (Jan. 2003), pp. 83–110. DOI: [10.1146/annurev.publhealth.24.100901.140927](https://doi.org/10.1146/annurev.publhealth.24.100901.140927).
- [580] Mayur Murali et al. “Ethnic minority representation in UK COVID-19 trials: systematic review and meta-analysis”. In: *BMC Medicine* 21.1 (Mar. 2023), p. 111. DOI: [10.1186/s12916-023-02809-7](https://doi.org/10.1186/s12916-023-02809-7).
- [581] Zara Kayani et al. “Reporting and representation of underserved groups in intervention studies for patients with multiple long-term conditions: a systematic review”. In: *Journal of the Royal Society of Medicine* 117.9 (Sept. 2024), pp. 302–317. DOI: [10.1177/01410768241233109](https://doi.org/10.1177/01410768241233109).
- [582] UK Kidney Association. *2021 NICE CKD Guideline*. 2021. URL: <https://ukkidney.org/renal-association/news/2021-nice-ckd-guideline>.
- [583] D Cormack and C Robson. *Classification and output of multiple ethnicities : considerations for monitoring Māori health*. Te Rōpū Rangahau Hauora a Eru Pōmare, 2010, p. 58. ISBN: 9780986468230.
- [584] ZL Rocha. “(Mixed) Racial formation in Aotearoa/New Zealand: framing biculturalism and ‘mixed race’ through categorisation”. In: *Kōtuitui: New Zealand Journal of Social Sciences Online* 7.1 (May 2012), pp. 1–13. DOI: [10.1080/1177083X.2012.670650](https://doi.org/10.1080/1177083X.2012.670650).
- [585] N Boven et al. “The implications of different ethnicity categorisation methods for understanding outcomes and developing policy in New Zealand”. In: *Kōtuitui: New Zealand Journal of Social Sciences Online* 15.1 (Jan. 2020), pp. 123–139. DOI: [10.1080/1177083X.2019.1657912](https://doi.org/10.1080/1177083X.2019.1657912).
- [586] ES Yao et al. “Classifying multiple ethnic identifications: Methodological effects on child, adolescent, and adult ethnic distributions”. In: *Demographic Research* 44 (Mar. 2021), pp. 481–512. DOI: [10.4054/DemRes.2021.44.21](https://doi.org/10.4054/DemRes.2021.44.21).
- [587] Ministry of Social Development New Zealand. *Improving how we report ethnicity*. URL: <https://www.msd.govt.nz/about-msd-and-our-work/tools/how-we-report-ethnicity.html>.

- [588] Tahu Kukutai and Paul Callister. *A "main" ethnic group? A "MAIN" ETHNIC GROUP? ETHNIC SELF-PRIORITISATION AMONG NEW ZEALAND YOUTH*. Tech. rep. 2009.
- [589] Statistics New Zealand. *Census results reflect Aotearoa New Zealand's diversity*. Oct. 2024. URL: <https://www.stats.govt.nz/news/census-results-reflect-aotearoa-new-zealands-diversity/>.
- [590] Matthew Croucher et al. "Dementia-Mate-Wareware-Action-Plan". In: (). URL: <https://www.tewhatuora.govt.nz/assets/For-the-health-sector/Specific-life-stage/Health-of-older-people/Dementia-Mate-Wareware-Action-Plan.pdf>.
- [591] M Dudley et al. "Mate wareware: Understanding 'dementia' from a Māori perspective". In: *New Zealand Medical Journal* 132.1503 (Oct. 2019), pp. 66–74.
- [592] Deloitte and Alzheimers New Zealand. *Dementia Economic Impact Report 2016*. Tech. rep. 2017, pp. 1–78. URL: <https://cdn.alzheimers.org.nz/wp-content/uploads/2021/05/Economic-Impacts-of-Dementia-2017.pdf>.
- [593] Alzheimers New Zealand and Deloitte. *Updated Dementia Economic Impact Report, 2011, New Zealand*. Tech. rep. Auckland, New Zealand, Jan. 2012. URL: https://cdn.alzheimers.org.nz/wp-content/uploads/2022/04/Updated_Dementia_Economic_Impact_Report_2012_New_Zealand.pdf.
- [594] Raimo Sulkava et al. "Prevalence of severe dementia in Finland". In: *Neurology* 35.7 (July 1985), pp. 1025–1025. DOI: [10.1212/WNL.35.7.1025](https://doi.org/10.1212/WNL.35.7.1025).
- [595] Silvan Licher et al. "Vitamin D and the Risk of Dementia: The Rotterdam Study". In: *Journal of Alzheimer's Disease* 60.3 (Oct. 2017), pp. 989–997. DOI: [10.3233/JAD-170407](https://doi.org/10.3233/JAD-170407).
- [596] Victoria Gil Martínez, Ana Avedillo Salas, and Sonia Santander Ballestín. "Vitamin Supplementation and Dementia: A Systematic Review". In: *Nutrients* 14.5 (Feb. 2022), p. 1033. DOI: [10.3390/nu14051033](https://doi.org/10.3390/nu14051033).
- [597] Lauren E. Mokry et al. "Genetically decreased vitamin D and risk of Alzheimer disease". In: *Neurology* 87.24 (Dec. 2016), pp. 2567–2574. DOI: [10.1212/WNL.0000000000003430](https://doi.org/10.1212/WNL.0000000000003430).
- [598] Wiki Open Street Map. *Scotland*. 2024. URL: <https://wiki.openstreetmap.org/wiki/Scotland>.
- [599] Wiki Open Street Map. *New Zealand*. 2023. URL: https://wiki.openstreetmap.org/wiki/New_Zealand.
- [600] A Black. *The IDI prototype spine's creation and coverage (Statistics New Zealand Working Paper No 16-03)*. No 16-03. Statistics New Zealand, 2016. ISBN: 9780908350544.
- [601] University of Auckland. *Growing Up in New Zealand research study*. 2023. URL: <https://www.growingup.co.nz/>.
- [602] Massey University. *New Zealand Health, Work and Retirement Study*. 2024. URL: <https://www.massey.ac.nz/about/colleges-schools-and-institutes/college-of-humanities-and-social-sciences/research-in-the-college-of-humanities-and-social-sciences/psychology-research/new-zealand-health-work-and-retirement-study-hwr/>.

- [603] University of Auckland. *LiLACS NZ*. 2024. URL: <https://www.fmhs.auckland.ac.nz/en/faculty/lilacs.html>.
- [604] Sue Wells et al. "Cohort Profile: The PREDICT Cardiovascular Disease Cohort in New Zealand Primary Care (PREDICT-CVD 19)". In: *International Journal of Epidemiology* (Dec. 2015), dyv312. DOI: [10.1093/ije/dyv312](https://doi.org/10.1093/ije/dyv312).
- [605] University of Otago. *The DUNEDIN Study*. 2024. URL: <https://dunedinstudy.otago.ac.nz/>.
- [606] Statistics New Zealand. *Statistics New Zealand Integrated Data Infrastructure*. URL: <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>.
- [607] Statistics New Zealand Tatauranga Aotearoa and New Zealand Government. *Integrated Data Infrastructure*. Aug. 2022. URL: <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>.
- [608] Statistics New Zealand. *Data in the IDI October 2024*. 2024. URL: <file:///C:/Users/s1577767/Downloads/October-2024-Data-in-the-IDI.pdf>.
- [609] Statistics New Zealand. *Data in the IDI*. URL: <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/data-in-the-idi#income>.
- [610] data.govt.nz. *Ngā Tikanga Paihere*. URL: <https://data.govt.nz/toolkit/data-ethics/nga-tikanga-paihere>.
- [611] Ngā Tikanga Paihere. *Guidelines 2020 Ngā Tikanga Paihere: a framework guiding ethical and culturally appropriate data use*. Tech. rep. URL: www.stats.govt.nz.
- [612] New Zealand Ministry of Health. *Current Data Access Policy*. URL: <https://www.health.govt.nz/publications/current-data-access-policy>.
- [613] New Zealand Government and New Zealand Ministry of Health. *Health and Disability Ethics Committees (HDEC)*. URL: <https://ethics.health.govt.nz/>.
- [614] Te Mana Raraunga. *Ngā mihi ki a koutou katoa*. URL: <https://www.temanararaunga.maori.nz/>.
- [615] Te Mana Raraunga. *Principles of Māori Data Sovereignty Definition of terms "He rei ngā niho, he paraoa ngā kauae"*. Tech. rep. URL: <https://www.un.org/development/desa/indigenouspeoples/>.
- [616] ANZSOG. *Overcoming preconceptions: How big data can gain a social licence in New Zealand*. URL: <https://anzsog.edu.au/news/overcoming-preconceptions/>.
- [617] Statistics New Zealand. *A social licence approach to trust*. 2018. URL: <https://www.stats.govt.nz/corporate/a-social-licence-approach-to-trust>.
- [618] Statistics New Zealand. *Stats NZ's social licence for data stewardship*. URL: <https://www.stats.govt.nz/corporate/stats-nzs-social-licence-for-data-stewardship>.
- [619] Te Kawa Mataaho Public Service Commission. *01 Kiwis count - Trust*. URL: <https://www.publicservice.govt.nz/research-and-data/kiwis-count-survey-2019/trust>.

- [620] Amy Hai Yan Chan et al. “Anti-dementia medication use in Aotearoa New Zealand: An exploratory study using health data from the Integrated Data Infrastructure (IDI)”. In: *Australian & New Zealand Journal of Psychiatry* 57.6 (June 2023), pp. 895–903. DOI: [10.1177/00048674221121091](https://doi.org/10.1177/00048674221121091).
- [621] BPAC New Zealand. *The pharmacological management of Alzheimer’s disease: The place of donepezil*. 2010. URL: <https://bpac.org.nz/bpj/2010/august/alzheimers.aspx>.
- [622] Dementia New Zealand. *Specific Dementia Medications*. URL: <https://dementia.nz/specific-dementia-medications/>.
- [623] June M. Tordoff et al. “Improvements in the prescribing of antipsychotics in dementia and psychogeriatric units in New Zealand”. In: *International Journal of Clinical Pharmacy* 38.4 (Aug. 2016), pp. 941–949. DOI: [10.1007/s11096-016-0318-1](https://doi.org/10.1007/s11096-016-0318-1).
- [624] IHACPA. *ICD-10-AM/ACHI/ACS*. URL: <https://www.ihacpa.gov.au/health-care/classification/icd-10-amachiacs>.
- [625] Jinfeng Zhao et al. “Constructing whole of population cohorts for health and social research using the New Zealand Integrated Data Infrastructure”. In: *Australian and New Zealand Journal of Public Health* 42.4 (Aug. 2018), pp. 382–388. DOI: [10.1111/1753-6405.12781](https://doi.org/10.1111/1753-6405.12781).
- [626] Sarah Cullum et al. “Do community-dwelling Māori and Pacific peoples present with dementia at a younger age and at a later stage compared with NZ Europeans?” In: *International Journal of Geriatric Psychiatry* 33.8 (Aug. 2018), pp. 1098–1104. DOI: [10.1002/gps.4898](https://doi.org/10.1002/gps.4898).
- [627] M Poland and Z Van Der Merwe. *VHIN: Using ACC claims data in the IDI*. 2019. URL: <https://vhin.co.nz/guides/using-acc-claims-data-in-the-idi/>.
- [628] Health New Zealand Te Whatu Ora. *National Minimum Dataset (hospital events)*. URL: <https://www.tewhatuora.govt.nz/for-health-professionals/data-and-statistics/nz-health-statistics/national-collections-and-surveys/collections/national-minimum-dataset-hospital-events>.
- [629] Institute for Health Metrics and Evaluation. *New Zealand National Minimum Dataset 2015*. 2015. URL: <https://ghdx.healthdata.org/record/new-zealand-national-minimum-dataset-2015>.
- [630] interRAI. *interRAI Worldwide*. URL: <https://interrai.org/>.
- [631] Statistics New Zealand Tatauranga Aotearoa. *How do we keep integrated data safe*. Aug. 2022. URL: <https://www.stats.govt.nz/integrated-data/how-we-keep-integrated-data-safe/>.
- [632] Statistics New Zealand. “Random Rounding”. 2016. URL: <https://data.govt.nz/assets/Uploads/data-confidentiality-random-rounding.pdf>.
- [633] Statistics New Zealand. *Meshblocks 2013*. URL: <https://datafinder.stats.govt.nz/layer/8347-meshblock-2013/>.
- [634] Statistics New Zealand. *Area Unit 2013*. URL: <https://datafinder.stats.govt.nz/layer/25743-area-unit-2013/>.

- [635] Jinfeng Zhao and Daniel J. Exeter. “Developing intermediate zones for analysing the social geography of Auckland, New Zealand”. In: *New Zealand Geographer* 72.1 (Apr. 2016), pp. 14–27. DOI: [10.1111/nzg.12110](https://doi.org/10.1111/nzg.12110).
- [636] University of Auckland. *Health Geography and Deprivation: Datazones 2013*. URL: <https://hgd.auckland.ac.nz/datazones/>.
- [637] Statistics New Zealand. *District Health Boards 2015*. URL: <https://datafinder.stats.govt.nz/layer/87883-district-health-board-2015/>.
- [638] Statistics New Zealand. *2013 Census*. URL: <https://www.stats.govt.nz/census/previous-censuses/2013-census/>.
- [639] Statistics New Zealand. *Territorial Authority 2013*. URL: <https://datafinder.stats.govt.nz/layer/25735-territorial-authority-2013/>.
- [640] Statistics New Zealand. *Statistical Area 1 2018 generalised*. URL: <https://datafinder.stats.govt.nz/layer/92210-statistical-area-1-2018-generalised/>.
- [641] Statistics New Zealand. *Statistical Area 2 2018 generalised*. URL: <https://datafinder.stats.govt.nz/layer/92212-statistical-area-2-2018-generalised/>.
- [642] Statistics New Zealand. *Meshblocks 2018 generalised*. URL: <https://datafinder.stats.govt.nz/layer/92197-meshblock-2018-generalised/>.
- [643] Daniel John Exeter et al. “The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand”. In: *PLOS ONE* 12.8 (Aug. 2017), e0181260. DOI: [10.1371/journal.pone.0181260](https://doi.org/10.1371/journal.pone.0181260).
- [644] DJ Exeter et al. “The Older Persons’ Index of Multiple Deprivation: Measuring the deprivation circumstances of older populations in Aotearoa New Zealand”. In: *Health & Place* 76 (July 2022), p. 102850. DOI: [10.1016/j.healthplace.2022.102850](https://doi.org/10.1016/j.healthplace.2022.102850).
- [645] Christian LoBue et al. “Traumatic brain injury history is associated with earlier age of onset of Alzheimer disease”. In: *The Clinical Neuropsychologist* 31.1 (Jan. 2017), pp. 85–98. DOI: [10.1080/13854046.2016.1257069](https://doi.org/10.1080/13854046.2016.1257069).
- [646] Sarah Cullum et al. “Predictors of mortality in Māori, Pacific Island, and European patients diagnosed with dementia at a New Zealand Memory Service”. In: *International Journal of Geriatric Psychiatry* 35.5 (May 2020), pp. 516–524. DOI: [10.1002/gps.5266](https://doi.org/10.1002/gps.5266).
- [647] Gary Cheung et al. “The Understanding and Experiences of Living with Dementia in Chinese New Zealanders”. In: *International Journal of Environmental Research and Public Health* 19.3 (Jan. 2022), p. 1280. DOI: [10.3390/ijerph19031280](https://doi.org/10.3390/ijerph19031280).
- [648] Adrian Martinez-Ruiz et al. “Living with Dementia in Aotearoa (LiDiA): a cross-sectional feasibility study protocol for a multiethnic dementia prevalence study in Aotearoa/New Zealand”. In: *BMJ Open* 11.5 (May 2021), e046143. DOI: [10.1136/bmjopen-2020-046143](https://doi.org/10.1136/bmjopen-2020-046143).
- [649] Adrian Martinez-Ruiz et al. “Diagnostic accuracy of 10/66 dementia protocol in Māori kaumātua (elders) living in Aotearoa New Zealand.” In: *The New Zealand medical journal* 135.1548 (Jan. 2022), pp. 42–53.

- [650] Adrian Martinez-Ruiz et al. “Diagnostic Accuracy of 10/66 Dementia Protocol in Fijian-Indian Elders Living in New Zealand”. In: *International Journal of Environmental Research and Public Health* 18.9 (May 2021), p. 4870. DOI: [10.3390/ijerph18094870](https://doi.org/10.3390/ijerph18094870).
- [651] Leah S. Richmond-Rakerd et al. “Associations of hospital-treated infections with subsequent dementia: nationwide 30-year analysis”. In: *Nature Aging* 4.6 (May 2024), pp. 783–790. DOI: [10.1038/s43587-024-00621-3](https://doi.org/10.1038/s43587-024-00621-3).
- [652] LS Richmond-Rakerd et al. “Longitudinal Associations of Mental Disorders With Dementia”. In: *JAMA Psychiatry* 79.4 (Apr. 2022), p. 333. DOI: [10.1001/jamapsychiatry.2021.4377](https://doi.org/10.1001/jamapsychiatry.2021.4377).
- [653] Lisa Underwood et al. “Long-term health conditions among household families in Aotearoa New Zealand: cross-sectional analysis of integrated Census and administrative data”. In: *New Zealand Medical Journal* 137.1596 (June 2024), pp. 20–34. DOI: [10.26635/6965.6370](https://doi.org/10.26635/6965.6370).
- [654] Daniel J. Exeter et al. “Developing SEP65: A Census-Derived Index of Socio-Economic Position Specifically for the Older Population in Aotearoa New Zealand”. In: *Social Indicators Research* 169.3 (Oct. 2023), pp. 973–991. DOI: [10.1007/s11205-023-03190-x](https://doi.org/10.1007/s11205-023-03190-x).
- [655] Sarah Ann Pinto et al. *Older people experiencing vulnerability and multiple disadvantage in New Zealand : a report on the needs of older people (65+) in health, housing, finance, social connection, and access*. ISBN: 9781991178541. URL: <https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/older-people-experiencing-vulnerability-and-multiple-disadvantage-in-new-zealand/older-people-vulnerability-and-multiple-disadvantage-june.pdf>.
- [656] Corina Grey et al. “Manawataki Fatu Fatu for ACCESS (Māori and Pacific Hearts in Unison for Achieving Cardiovascular Care in Equity StudieS). Protocol for a Mixed Methods Programme of Research”. In: *International Journal of Qualitative Methods* 22 (Oct. 2023). DOI: [10.1177/16094069231176348](https://doi.org/10.1177/16094069231176348).
- [657] Heart Foundation New Zealand. *Cardiovascular Disease Risk Assessment and Management*. URL: <https://www.heartfoundation.org.nz/professionals/health-professionals/cvd-consensus-summary>.
- [658] M Crooke. “New Zealand cardiovascular guidelines: best practice evidence-based guideline: the assessment and management of cardiovascular risk December 2003.” In: *The Clinical biochemist. Reviews* 28.1 (Feb. 2007), pp. 19–29.
- [659] KM Brewer et al. “What are the gaps in cardiovascular risk assessment and management in primary care for Māori and Pacific people in Aotearoa New Zealand? Protocol for a systematic review.” In: *BMJ open* 12.6 (June 2022), e060145. DOI: [10.1136/bmjopen-2021-060145](https://doi.org/10.1136/bmjopen-2021-060145).
- [660] Briar Warin et al. “Geography matters: the prevalence of diabetes in the Auckland Region by age, gender and ethnicity”. In: *The New Zealand medical journal* 129.1436 (June 2016), pp. 25–37.
- [661] Health New Zealand Te Whatu Ora. *Dementia Mate Wareware*. URL: <https://www.tewhatauora.govt.nz/for-health-professionals/clinical-guidance/specific-life-stage-health-information/health-of-older-people/dementia-mate-wareware>.

- [662] Office for National Statistics. “A_Beginners_Guide_to_UK_Geography_2023”. In: (). URL: <https://geoportal.statistics.gov.uk/datasets/c0db0e8c67d04935bcf1749ca/about>.
- [663] Scottish Government, Geographic Information Science & Analysis Team, and Rural and Environment Science and Analytical Services Division. *Scottish Government Urban Rural Classification 2020*. Tech. rep. May 2022. URL: <https://www.gov.scot/binaries/content/documents/govscot/publications/advice-and-guidance/2022/05/scottish-government-urban-rural-classification-2020/documents/scottish-government-urban-rural-classification-2020/govscot%3Adocument/scottish-government-urban-rural-classification-2020.pdf>.
- [664] National Records of Scotland (NRS). *Scottish Postcode Directory (SPD)*. URL: <https://www.nrscotland.gov.uk/statistics-and-data/geography/nrs-postcode-extract>.
- [665] Scottish Government. *Small area statistics*. URL: <https://www.gov.scot/collections/small-area-statistics/>.
- [666] V Carstairs and R Morris. “Deprivation and Health in Scotland”. In: *Aberdeen University Press* (1991).
- [667] Scottish Government. *Scottish Index of Multiple Deprivation 2020*. URL: <https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/>.
- [668] Public Health Scotland. *Deprivation guidance for analysts*. Tech. rep. Dec. 2023. URL: www.publichealthscotland.scot.
- [669] Public Health Scotland. *Electronic Data Research and Innovation Service (eDRIS): National Safe Haven (NSH)*. URL: <https://publichealthscotland.scot/services/data-research-and-innovation-services/electronic-data-research-and-innovation-service-edris/national-safe-haven-nsh/>.
- [670] The Scottish Government. *Estimated and Projected Diagnosis Rates for Dementia in Scotland: 2014-2020*. Tech. rep. 2016.
- [671] Global Council on Brain Health. *Building Better Brain Health for All People: GCBH Recommendations on Removing Barriers and Improving Opportunities Around the World*. Tech. rep. Washington, DC: Global Council on Brain Health, June 2023. DOI: [10.26419/pia.00107.001](https://doi.org/10.26419/pia.00107.001).
- [672] Nicolas Farina et al. “Perceptions of dementia amongst the general public across Latin America: a systematic review”. In: *Aging & Mental Health* 25.5 (May 2021), pp. 787–796. DOI: [10.1080/13607863.2020.1725738](https://doi.org/10.1080/13607863.2020.1725738).
- [673] IJ Deary, L Whalley, and JM Starr. “A lifetime of intelligence: follow-up studies of the Scottish Mental Surveys of 1932 and 1947”. In: Washington, DC: American Psychological Association, 2009. DOI: <https://psycnet.apa.org/doi/10.1037/11857-000>.
- [674] Adele M Taylor, Alison Pattie, and Ian J Deary. “Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936”. In: *International Journal of Epidemiology* 47.4 (Aug. 2018), pp. 1042–1042. DOI: [10.1093/ije/dyy022](https://doi.org/10.1093/ije/dyy022).

- [675] AM Nuttall et al. “Maternity records in Edinburgh and Aberdeen in 1936: a comparison”. In: *The Journal of the Royal College of Physicians of Edinburgh* 45.1 (2015), pp. 76–83. DOI: [10.4997/JRCPE.2015.117](https://doi.org/10.4997/JRCPE.2015.117).
- [676] Markus Jokela et al. “Low Childhood IQ and Early Adult Mortality: The Role of Explanatory Factors in the 1958 British Birth Cohort”. In: *Pediatrics* 124.3 (Sept. 2009), e380–e388. DOI: [10.1542/peds.2009-0334](https://doi.org/10.1542/peds.2009-0334).
- [677] Iva Čukić et al. “Childhood IQ and survival to 79: Follow-up of 94% of the Scottish Mental Survey 1947”. In: *Intelligence* 63 (July 2017), pp. 45–50. DOI: [10.1016/j.intell.2017.05.002](https://doi.org/10.1016/j.intell.2017.05.002).
- [678] AHM Kilgour, JM Starr, and LJ Whalley. “Associations between childhood intelligence (IQ), adult morbidity and mortality”. In: *Maturitas* 65.2 (Feb. 2010), pp. 98–105. DOI: [10.1016/j.maturitas.2009.09.021](https://doi.org/10.1016/j.maturitas.2009.09.021).
- [679] CM Calvin et al. “Childhood intelligence in relation to major causes of death in 68 year follow-up: prospective population study”. In: *BMJ* (June 2017), j2708. DOI: [10.1136/bmj.j2708](https://doi.org/10.1136/bmj.j2708).
- [680] CM Calvin et al. “Intelligence in youth and all-cause-mortality: systematic review with meta-analysis”. In: *International Journal of Epidemiology* 40.3 (June 2011), pp. 626–644. DOI: [10.1093/ije/dyq190](https://doi.org/10.1093/ije/dyq190).
- [681] TC Russ et al. “Childhood Cognitive Ability and Incident Dementia”. In: *Epidemiology* 28.3 (May 2017), pp. 361–364. DOI: [10.1097/EDE.0000000000000626](https://doi.org/10.1097/EDE.0000000000000626).
- [682] CE Brett and IJ Deary. “Realising health data linkage from a researcher’s perspective: Following up the 6-day sample of the Scottish mental survey 1947”. In: *Longitudinal and Life Course Studies* 5.3 (2014), pp. 283–298. DOI: [10.14301/llcs.v5i3.266](https://doi.org/10.14301/llcs.v5i3.266).
- [683] Public Health Scotland. *Data and intelligence Previously ISD Scotland*. Jan. 2024. URL: <https://www.isdscotland.org/>.
- [684] Public Health Scotland. *The Public Benefit and Privacy Panel for Health and Social Care – Why does it exist?* URL: <https://www.informationgovernance.scot.nhs.uk/pbpphsc/test-phase-1-pbpp-why-does-it-exist/>.
- [685] National Records of Scotland. *Scottish Longitudinal Study (SLS)*. URL: <https://www.nrscotland.gov.uk/statistics-and-data/scottish-longitudinal-study-sls>.
- [686] Samantha Alvarez-Madrazo et al. “Data Resource Profile: The Scottish National Prescribing Information System (PIS)”. In: *International Journal of Epidemiology* 45.3 (June 2016), pp. 714–715. DOI: [10.1093/ije/dyw060](https://doi.org/10.1093/ije/dyw060).
- [687] T Wilkinson. *Dementia risk prediction model - definitions of variables, [dataset]*. 2020. DOI: <https://doi.org/10.7488/ds/2823>.
- [688] NHS Digital. *NHS Data Model and Dictionary: Dementia*. URL: <https://archive.datadictionary.nhs.uk/DD%20Release%20March%202021/Covid19PRA/Dementia.html>.
- [689] V Kuan et al. “A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service”. In: *The Lancet Digital Health* 1.2 (June 2019), e63–e77. DOI: [10.1016/S2589-7500\(19\)30012-3](https://doi.org/10.1016/S2589-7500(19)30012-3).

- [690] V Kuan et al. *PH148 / 296 - Dementia. Phenotype Library [Online]*. Oct. 2021. URL: <http://phenotypes.healthdatagateway.org/phenotypes/PH148/version/296/detail/>.
- [691] Matthew H. Iveson et al. "Childhood cognitive ability and self-harm and suicide in later life". In: *SSM - Population Health* 25 (Mar. 2024), p. 101592. DOI: [10.1016/j.ssmph.2023.101592](https://doi.org/10.1016/j.ssmph.2023.101592).
- [692] EL Ball et al. "Childhood intelligence and risk of depression in later-life: A longitudinal data-linkage study". In: *SSM - Population Health* 25 (Mar. 2024), p. 101560. DOI: [10.1016/j.ssmph.2023.101560](https://doi.org/10.1016/j.ssmph.2023.101560).
- [693] Jianping Jia et al. "The prevalence of dementia in urban and rural areas of China". In: *Alzheimer's & Dementia* 10.1 (Jan. 2014), pp. 1–9. DOI: [10.1016/j.jalz.2013.01.012](https://doi.org/10.1016/j.jalz.2013.01.012).
- [694] H Chen et al. "Cohort Profile: The ONtario Population Health and Environment Cohort (ONPHEC)". In: *International Journal of Epidemiology* (Apr. 2016), dyw030. DOI: [10.1093/ije/dyw030](https://doi.org/10.1093/ije/dyw030).
- [695] Geneviève Arsenault-Lapierre et al. "Rural and urban differences in quality of dementia care of persons with dementia and caregivers across all domains: a systematic review". In: *BMC Health Services Research* 23.1 (Jan. 2023), p. 102. DOI: [10.1186/s12913-023-09100-8](https://doi.org/10.1186/s12913-023-09100-8).
- [696] SH Rhew et al. "Rural health disparities in health care utilization for dementia in Minnesota". In: *The Journal of Rural Health* 39.3 (June 2023), pp. 656–665. DOI: [10.1111/jrh.12700](https://doi.org/10.1111/jrh.12700).
- [697] DJ Exeter et al. "The geographic harmonisation of Scotland's small area census data, 1981 to 2011". In: *Health & Place* 57 (May 2019), pp. 22–26. DOI: [10.1016/j.healthplace.2019.02.003](https://doi.org/10.1016/j.healthplace.2019.02.003).
- [698] S Toot et al. "Causes of Hospital Admission for People With Dementia: A Systematic Review and Meta-Analysis". In: *Journal of the American Medical Directors Association* 14.7 (July 2013), pp. 463–470. DOI: [10.1016/j.jamda.2013.01.011](https://doi.org/10.1016/j.jamda.2013.01.011).
- [699] R. *CRAN Mirrors*. URL: <https://cran.r-project.org/mirrors.html>.
- [700] NHS England. *Synnovis cyber attack – statement from NHS England*. June 2024. URL: <https://www.england.nhs.uk/2024/06/synnovis-cyber-attack-statement-from-nhs-england/>.
- [701] UK Government, National Data Guardian (NDG), and NHS England. *NDG and NHS England issue joint statement about NHS Data Security and Protection Toolkit*. Sept. 2024. URL: <https://www.gov.uk/government/news/ndg-and-nhs-england-issue-joint-statement-about-nhs-data-security-and-protection-toolkit>.
- [702] NHS England. *Update on cyber incident: clinical impact in South East London – Friday 14 June 2024*. June 2024. URL: <https://www.england.nhs.uk/london/2024/06/14/update-on-cyber-incident-clinical-impact-in-south-east-london-friday-14-june-2024/>.
- [703] NHS Scotland. *Cyber attacks on NHS Dumfries and Galloway*. 2024. URL: <https://www.nhs.uk/scotland/cyberattacks/>.

- [704] P Carter, GT Laurie, and M Dixon-Woods. “The social licence for research: Why care.data ran into trouble”. In: *Journal of Medical Ethics* 41.5 (2015), pp. 404–409. DOI: [10.1136/medethics-2014-102374](https://doi.org/10.1136/medethics-2014-102374).
- [705] M Pollok, E Gray, and J Sakali. *Public perspectives on the use of health data in research and innovation*. Tech. rep. Ipsos Scotland, May 2022. URL: <https://ipsos.uk/terms..>
- [706] Statistics New Zealand. *New Zealanders’ trust in key institutions declines*. Sept. 2024. URL: <https://www.stats.govt.nz/news/new-zealanders-trust-in-key-institutions-declines/>.
- [707] Gary Cheung et al. “Dementia prevalence estimation among the main ethnic groups in New Zealand: a population-based descriptive study of routinely collected health data”. In: *BMJ Open* 12.9 (Sept. 2022), e062304. DOI: [10.1136/bmjopen-2022-062304](https://doi.org/10.1136/bmjopen-2022-062304).
- [708] E. Ma’u et al. “Estimating the Prevalence of Dementia in New Zealand Using Capture-Recapture Analysis on Routinely Collected Health Data”. In: *International Journal of Geriatric Psychiatry* 39.8 (Aug. 2024). DOI: [10.1002/gps.6131](https://doi.org/10.1002/gps.6131).
- [709] European Environment Agency. *Degree of Urbanisation (DEGURBA)*. 2017. URL: <https://www.eea.europa.eu/data-and-maps/data/external/degree-of-urbanisation-degurba>.
- [710] “GHSL_Data_Package_2023”. In: ().
- [711] *GEO Human Planet*. URL: <https://earthobservations.org/organization/work-programme/geo-human-planet>.
- [712] “geo_work_programme_2023_2025_summary_document_v3_20221214”. In: ().
- [713] World Bank. *How do we define cities, towns and rural areas?* 2020. URL: <https://blogs.worldbank.org/en/sustainablecities/how-do-we-define-cities-towns-and-rural-areas>.
- [714] Office for National Statistics. “A Beginners Guide to UK Geography 2023: Geographical areas frequently used in the publication of official statistics”. In: (2023). URL: <https://geoportal.statistics.gov.uk/datasets/c0db0e8c67d04935bcf1749ca6027fef/about>.
- [715] Northern Ireland Statistics and Research Agency (NISRA). *New statistical output geographies for Northern Ireland derived from Census 2021*. Tech. rep. 2023. URL: [New%20statistical%20output%20geographies%20for%20Northern%20Ireland%20derived%20from%20Census%202021](https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/new-statistical-output-geographies-for-northern-ireland-derived-from-census-2021).
- [716] Northern Ireland Statistics and Research Agency (NISRA). *Review of the Statistical Classification and Delineation of Settlements*. Tech. rep. 2015. URL: https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/review-of-the-statistical-classification-and-delineation-of-settlements-march-2015_0.pdf.
- [717] Office for National Statistics. *Census 2021 geographies*. URL: <https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies>.

- [718] Public Health Scotland. *Geography, population and deprivation support*. Mar. 2024. URL: <https://publichealthscotland.scot/services/geography-population-and-deprivation-support/deprivation/carstairs-index/>.
- [719] Etuini Ma'u et al. "Differences in the potential for dementia prevention between major ethnic groups within one country: A cross sectional analysis of population attributable fraction of potentially modifiable risk factors in New Zealand". In: *The Lancet Regional Health - Western Pacific* 13 (Aug. 2021), p. 100191. DOI: [10.1016/j.lanwpc.2021.100191](https://doi.org/10.1016/j.lanwpc.2021.100191).
- [720] Wenbo Wu et al. "Natural language processing to identify social determinants of health in Alzheimer's disease and related dementia from electronic health records". In: *Health Services Research* 58.6 (Dec. 2023), pp. 1292–1302. DOI: [10.1111/1475-6773.14210](https://doi.org/10.1111/1475-6773.14210).
- [721] Laura C. MacLagan et al. "Can Patients with Dementia Be Identified in Primary Care Electronic Medical Records Using Natural Language Processing?" In: *Journal of Healthcare Informatics Research* 7.1 (Mar. 2023), pp. 42–58. DOI: [10.1007/s41666-023-00125-6](https://doi.org/10.1007/s41666-023-00125-6).
- [722] Ivan Koychev et al. "Identification of preclinical dementia according to ATN classification for stratified trial recruitment: A machine learning approach". In: *PLOS ONE* 18.10 (Oct. 2023), e0288039. DOI: [10.1371/journal.pone.0288039](https://doi.org/10.1371/journal.pone.0288039).
- [723] The Euan Macdonald Centre and University of Edinburgh. *MND Register*. URL: <https://www.euanmacdonaldcentre.org/research/research-case-studies/mnd-register>.
- [724] Patrick K. A. Kearns et al. "Regional variation in the incidence rate and sex ratio of multiple sclerosis in Scotland 2010–2017: findings from the Scottish Multiple Sclerosis Register". In: *Journal of Neurology* 266.10 (Oct. 2019), pp. 2376–2386. DOI: [10.1007/s00415-019-09413-x](https://doi.org/10.1007/s00415-019-09413-x).
- [725] Rob Baxter et al. "The Scottish Medical Imaging Archive: 57.3 Million Radiology Studies Linked to Their Medical Records". In: *Radiology: Artificial Intelligence* 6.1 (Jan. 2024). DOI: [10.1148/ryai.220266](https://doi.org/10.1148/ryai.220266).
- [726] *NRS Neuroprogressive and Dementia Clinical Research Network*. Tech. rep. 2022. URL: www.nrs.org.uk/ndn.
- [727] *Privacy Information Notice: How NRS Neuroprogressive and Dementia use your information*. Tech. rep.
- [728] Romana Pylypchuk et al. "Cardiovascular disease risk prediction equations in 400000 primary care patients in New Zealand: a derivation and validation study". In: *The Lancet* 391.10133 (May 2018), pp. 1897–1907. DOI: [10.1016/S0140-6736\(18\)30664-0](https://doi.org/10.1016/S0140-6736(18)30664-0).
- [729] JK Zakarias et al. "Geographical Variation in the Diagnostic Rate and Quality of Dementia Diagnoses". In: *Journal of Alzheimer's Disease* 69.2 (May 2019), pp. 513–520. DOI: [10.3233/JAD-190030](https://doi.org/10.3233/JAD-190030).
- [730] Rita V. Krishnamurthi et al. "Lived Experience of Dementia in the New Zealand Indian Community: A Qualitative Study with Family Care Givers and People Living with Dementia". In: *International Journal of Environmental Research and Public Health* 19.3 (Jan. 2022), p. 1432. DOI: [10.3390/ijerph19031432](https://doi.org/10.3390/ijerph19031432).

- [731] J Hazan et al. “Has COVID-19 affected dementia diagnosis rates in England?” In: *International Journal of Geriatric Psychiatry* 38.7 (July 2023). DOI: [10.1002/gps.5976](https://doi.org/10.1002/gps.5976).
- [732] R Canaway et al. “Undercounting diagnoses in Australian general practice: a data quality study with implications for population health reporting”. In: *BMC Medical Informatics and Decision Making* 24.1 (June 2024), p. 155. DOI: [10.1186/s12911-024-02560-w](https://doi.org/10.1186/s12911-024-02560-w).
- [733] David S. Jones and Jeremy A. Greene. “Is Dementia in Decline? Historical Trends and Future Trajectories”. In: *New England Journal of Medicine* 374.6 (Feb. 2016), pp. 507–509. DOI: [10.1056/NEJMp1514434](https://doi.org/10.1056/NEJMp1514434).
- [734] National Records of Scotland. *Scotland’s population grows*. Oct. 2024. URL: <https://www.nrscotland.gov.uk/news/2024/scotlands-population-grows>.
- [735] W Allen. *Geographical Distribution and Characteristics of Long-Term International Migration Flows to Scotland*. Tech. rep. Oxford: Migration Observatory briefing, COMPAS, University of Oxford, Sept. 2013. URL: www.migrationobservatory.ox.ac.uk.
- [736] National Records of Scotland. *Age-standardised death rates calculated using the European Standard Population-Methodology*. Tech. rep. URL: <https://www.nrscotland.gov.uk/files//statistics/age-standardised-death-rates-esp/2017/age-standardised-17-methodology.pdf>.
- [737] New Zealand Ministry of Health. *Position paper on Māori health analytics - age standardisation*. Tech. rep. New Zealand Ministry of Health, Nov. 2018, pp. 1–11. URL: https://www.health.govt.nz/system/files/2018-11/position-paper-maori-health-analytics-age-standardisation-nov18_0.pdf.
- [738] Bridget Robson et al. “Age standardisation – an indigenous standard?” In: *Emerging Themes in Epidemiology* 4.1 (Dec. 2007), p. 3. DOI: [10.1186/1742-7622-4-3](https://doi.org/10.1186/1742-7622-4-3).
- [739] Australian Institute of Health and Welfare. *Principles on the use of direct age-standardisation in administrative data collections For measuring the gap between Indigenous and non-Indigenous Australians*. Australian Institute of Health and Welfare, 2011. ISBN: 9781742492148. URL: <https://www.aihw.gov.au/reports/indigenous-australians/principles-on-the-use-of-direct-age-standardisation/contents/table-of-contents>.
- [740] Vahid Payman, Susan Yates, and Sarah Cullum. “Early onset dementia in New Zealand Pacific boxers: a case series”. In: *The New Zealand medical journal* 131.1474 (May 2018), pp. 20–26.
- [741] Y Marcon et al. “Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD”. In: *PLOS Computational Biology* 17.3 (Mar. 2021), e1008880. DOI: [10.1371/journal.pcbi.1008880](https://doi.org/10.1371/journal.pcbi.1008880).
- [742] D Doiron et al. “Software Application Profile: Opal and Mica: open-source software solutions for epidemiological data management, harmonization and dissemination”. In: *International Journal of Epidemiology* 46.5 (Oct. 2017), pp. 1372–1378. DOI: [10.1093/ije/dyx180](https://doi.org/10.1093/ije/dyx180).

- [743] A Gaye et al. “DataSHIELD: taking the analysis to the data, not the data to the analysis”. In: *International Journal of Epidemiology* 43.6 (Dec. 2014), pp. 1929–1944. DOI: [10.1093/ije/dyu188](https://doi.org/10.1093/ije/dyu188).
- [744] S Lenz, M Hess, and H Binder. “Deep generative models in DataSHIELD”. In: *BMC Medical Research Methodology* 21.1 (Dec. 2021), p. 64. DOI: [10.1186/s12874-021-01237-6](https://doi.org/10.1186/s12874-021-01237-6).
- [745] DataSHIELD Secure Bioscience Collaboration. *DataSHIELD: A software for secure data science collaboration*. URL: <https://datashield.org/>.
- [746] Isabelle Budin-Ljøsne et al. “DataSHIELD: An Ethically Robust Solution to Multiple-Site Individual-Level Data Analysis”. In: *Public Health Genomics* 18.2 (2015), pp. 87–96. DOI: [10.1159/000368959](https://doi.org/10.1159/000368959).
- [747] Università Degli Studi Di Verona. *ORCHESTRA: Connecting European cohorts*. 2020. URL: <https://orchestra-cohort.eu/>.
- [748] European Commission. *European Health Data Space*. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en.
- [749] Cristina Menni et al. “Real-time tracking of self-reported symptoms to predict potential COVID-19”. In: *Nature Medicine* 26.7 (July 2020), pp. 1037–1040. DOI: [10.1038/s41591-020-0916-2](https://doi.org/10.1038/s41591-020-0916-2).
- [750] Alzheimer Europe. *Data sharing in dementia research - the EU landscape*. URL: https://www.alzheimer-europe.org/sites/default/files/2021-10/Data%20sharing%20in%20dementia%20research%20-%20the%20EU%20landscape_Alzheimer%20Europe.pdf.
- [751] Equator Network. *Equator Network: Enhancing the QUALity and Transparency Of health Research. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement*. URL: <https://www.equator-network.org/reporting-guidelines/record/>.