Social Media Narratives in Non-Communicable Diseases – Their Dynamics and Value for Patients, Communities and Health Researchers

Joanna Taylor

Thesis submitted for the degree of

Doctor of Philosophy
(Global Health)

University of Edinburgh
2020

Declaration

This thesis represents my original work and has not been submitted for any other degree or professional qualification.

Joanna Taylor

Edinburgh, United Kingdom

27th September 2020

Acknowledgements

I greatly appreciate the support and guidance that my supervisor, Dr Claudia Pagliari, has given me over the past six years. She has patiently and persistently helped to instill academic rigour into my research, provided intellectual challenge, introduced me to new concepts and members of the research community and helped me navigate the requirements of a PhD, as well as the expectations of journal editors. Her encouragement and advice were invaluable as I strove to complete this research whilst fulfilling my professional and personal commitments. Without her support, I would not have been awarded The Edinburgh Award for Professionalism in Community Engagement.

My sincerest appreciation also go to Dr Miles Osborne (co-supervisor, 2013-2015), who provided early guidance on the informatics elements of the research and in the pilot study, as well as to Dr Dave Murray-Rust (review panel member, 2014-2017 and co-supervisor, 2017-2020) and Professor Liz Grant (review panel member, 2014-2020), for their annual feedback on my progress and the outcome of my research.

The studies and subsequent publications did not occur in isolation. I would therefore like to thank: Marshall Dozier for her advice on the search strategy of the scoping review; Dr Santosh Vijaykumar for his feedback on an earlier version of the protocol; the EPPI-Centre, at the UCL Institute of Education; and Symplur for the use of EPPI Reviewer and the Symplur Transcript and Analytics tools, respectively. When conducting the empirical studies, I was cognizant that real people were creating the data collected and therefore I am also thankful to Chris Pointon, Kate Granger's widower, who agreed to the longitudinal illness trajectory study (Chapter 8) and the publication of its results, as well as all those, including Kate Granger herself, for posting about their patient experiences on social media.

Thanks to the editors and manuscript reviewers from *Research Ethics*, *BMJ Open*, *BMC Palliative Care* and *npj Digital Medicine*, as well as the conference organizers and members of the research community from the Stanford Medicine X conference, Medicine 2.0 Summit and World Congress, the International Conference on Global Health Challenges, the World Cancer Congress and the British Computing Society Health Informatics Scotland Conference for the opportunity to share my research and for the interesting discussions and feedback.

Although my employer, Ernst and Young, played no role in conducting the research, I would like to acknowledge the financial support that I received from the firm to attend and present at international conferences, as well as the ongoing encouragement from my colleagues and clients. Without their support and contributions to my professional and academic development, the International Forum for the Advancement of Health would not have recognised me as a Top 100 Healthcare Leader in 2019.

Finally, my decision and ability to further my academic development through this PhD, whilst working full-time and raising a young family, would not have been possible without my family. I wish to acknowledge the support and love of my husband Peter, my parents John and Norah and my daughters Emily and Isabelle.

Executive Summary

Background: Usage of social media is now widespread and growing, as is the number of people living with Non-Communicable Diseases (NCDs) such as diabetes and cancer. This thesis examines how social media are being used to share or discuss NCDs and the benefits, challenges and implications of these trends as a manifestation of digital public health.

Aim and research questions: The aim of this research is to address the gap in empirical, evidence-based research into the secondary use of data from social media to understand patient health issues and inform public health research into NCDs. To this end, seven research questions, each linked to a sub-project, were defined and tested during the course of the six-year programme:

1. What is the status of **the existing multi-disciplinary research literature** based on analysis of data posted on social media for public health research, and where are the gaps in this research?
2. Can existing **systematic review methods** be re-purposed and applied to analyse data posted on social media?
3. How are research sponsors and researchers addressing the **ethical challenges** of analysing data posted on social media?
4. To what extent are **diabetes-related posts on Twitter** relevant to the clinical condition and what topics and intentions are represented in these posts?
5. In what ways do people affected by **Type 1 diabetes** use *different* social media (e.g. for social interaction, support-seeking, information-sharing) and what are the implications for researchers wishing to use these data sources in their studies?
6. Are these differences in platform usage and associated data types also seen in people affected by **lung cancer**?
7. Can characteristic illness **trajectories** be seen in a cancer patient's digital narrative and what insights can be gained to inform palliative care services?

Methods: A range of different qualitative and quantitative methods and frameworks were used to address each of the research questions listed. Arksey and O'Malley's five-stage scoping review framework and the PRISMA guidelines are applied to the systematic scoping review of existing literature. The PRISMA guidelines and checklist are re-purposed and applied to the manual extraction and analysis of social media posts. Bjerglund-Andersen and Söderqvist's typology of social media uses in research and

Conway's taxonomy of ethical considerations are used to classify the ethics guidelines available to researchers. The findings of these were used to inform the research design of the four empirical studies.

The methods applied in the conduct of the empirical studies include a content and narrative analysis of cross-sectional and longitudinal data sourced from Twitter, Facebook, the Type 1 diabetes discussion forum on Diabetes.co.uk and the lung cancer discussion forum on Macmillan.org.uk, as well as the application of Bales' Interaction Process Analysis and Emanuel and Emanuel's framework for a good death.

Results: Of the 49 systematic, quasi-systematic and scoping reviews identified, 24 relate to the secondary use of data from social media, with eight of these focused on infectious disease surveillance and only two on NCDs. Existing reviews tend to be fragmented, narrow in scope and siloed in different academic communities, with limited consideration of the different types of data, analytical methods and ethical issues involved, therefore creating a need for further reviews to synthesise the emerging evidence-base.

The rapid increase in the volume of published research is evident, from the results of RQ1, with 87% of the eligible studies published between 2013-2017. Of the 105 eligible empirical studies that focused on NCDs, cancer (54%) and diabetes (20%) dominate the literature. Data is sourced from Twitter (26%), Facebook (14%) and blogs (10%), conducted, published and funded by the medical community. Since 2012, automated methods have increasingly been applied to extract and analyse large volumes of data.

Those that use manual methods for extraction did not apply a consistent approach to doing so; the PRISMA guidelines and checklist were therefore re-purposed and applied to analyse data extracted from social media in response to RQ2. The deficit of ethical guidance available to inform research that involves social media data was also identified as a result of RQ3 and the guidelines provided by the ESRC, BPS, AoIR and NIHR were prioritised for the purposes of this research project.

Results from the four empirical studies (RQ4-7) reveal that different forms of social interaction and support are represented in the variety of social media platforms available and that this is influenced by the type and nature of the condition with which people are affected, as well as the affordances offered by such platforms.

In the pilot study associated with RQ4, Twitter was identified as a 'noisy' source of data about diabetes, with only 66% of the sample being relevant to the clinical condition. Twelve per cent of the eligible sample was associated with Type 2 diabetes, compared to 6% for Type 1, and most were information-giving in nature (49%) and correlated with the diagnosis, treatment and management of the condition (44%).

A comparison of Twitter to the Type 1 Diabetes community on Facebook and the discussion forum on Diabetes.co.uk for RQ5 indicated that all three social media platforms were used to disseminate information about the condition. However, the Type 1 Diabetes Group on Facebook and the Type 1 discussion forum on Diabetes.co.uk were also used for social interaction and peer support, hence defying the generalisations made in public health studies, where social media platforms were often considered equal or synonymous.

The results from the third empirical study into lung cancer (RQ6) support this, indicating that, by virtue of their digital architecture, user base and self-moderating communities, the Lung Cancer Support Group on Facebook and the lung cancer discussion forum on Macmillan.org.uk are more successful in their utility for social interaction and emotional and informational support. Meanwhile, the sample derived from Twitter hashtags showed greater companionship support.

The final empirical study in this PhD research project is associated with RQ7 and used longitudinal data posted by a terminally ill patient on Twitter. This revealed that patient activity on social media mirrors the different phases of the end-of-life illness trajectory described in the literature and that it is comparable to or compliments insights garnered using more traditional qualitative research techniques. It also shows the value of such innovative methods for understanding how terminal disease is experienced by and affects individuals, how they cope, how support is sought and obtained and how patients feel about the ability of palliative care services to meet their needs at different stages.

Conclusions: The analysis of health data posted on social media continues to be an expanding and evolving field of multi-disciplinary research. The results of the studies included in this thesis reveal the emergence of new methods and ethical considerations to inform research design as well as ethics policy. The re-purposed PRISMA guidelines and checklist were presented at the 2014 Medicine 2.0 Summit and World Congress whilst the review of ethical guidelines was published in the *Research Ethics* journal.

The four empirical studies that extracted and analysed data from social media provide novel insight into the social narratives of those impacted by diabetes and cancer and can be used to inform future research and practice. The results of these studies have, to date, been presented at four international conferences and published in *npj Digital Medicine* and *BMC Palliative Care*.

Although this thesis and associated publications contribute to an emerging body of knowledge, further research is warranted into the manual versus automated techniques that can be applied and the differences in social interaction and support needed by people affected by different NCDs.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

AoIR	Association of Internet Researchers
ACM	Association of Computer Machinery
AHRC	Arts and Humanities Research Council
API	Application Programming Interface
ASSIA	Applied Social Sciences Index and Abstracts
BBSRC	Biotechnology and Biological Sciences Research Council
BPS	The British Psychological Society
EPSRC	Engineering and Physical Sciences Research Council
ESRC	Economic and Social Research Council
ERCP	Endoscopic Retrograde Cholangiopancreatographies
GOARN	Global Outbreak Alert and Response Network
GPS	Global Positioning System
HCP	Healthcare Professional
HIV/AIDS	Human Immunodeficiency Virus / Acquired Immunodeficiency Syndrome
H7N9	Avian influenza
IEEE	Institute of Electrical and Electronics Engineers
IMR	Internet Mediated Research
IPA	Interaction Process Analysis
IRB	Institutional Review Board
LCHF	Low Carbohydrate, High Fat diet
MeSH	Medical Subject Headings
MBE	Member of the Order of the British Empire
MERS	Middle East respiratory syndrome
MRC	Medical Research Council
NCD	Non-Communicable Disease
NERC	Natural Environment Research Council
NGO	Non-Governmental Organisation
NHS	National Health Service (United Kingdom)
NIH	National Institutes of Health (USA)
NIHR	National Institute of Health Research (United Kingdom)
NLP	Natural Language Processing
NSCLC	Non-Small Cell Lung Cancer
PRISMA	Preferred Reporting Items of Systematic Reviews and Meta-Analyses
RCUK	Research Councils United Kingdom
RQ	Research Question
SCLC	Small Cell Lung Cancer
STD	Sexually Transmitted Disease
STI	Sexually Transmitted Infection
STFC	Science and Technology Facilities Council
TB	Tuberculosis
T1D	Type 1 Diabetes
URL	Uniform Resource Locator
UTC	Universal Time Coordinated
WHO	World Health Organization

1. Introduction

The use of social media is widely accepted to be on the increase. When combined with the increasing number of people living with Non-Communicable Diseases (NCDs), this thesis provides the opportunity to evaluate and understand the potential benefits, challenges and implications of these trends in the context of health research.

This chapter provides an introduction to social media in the context of eHealth and the use of social media in health research and as a source of data for conducting empirical research and surveillance. It also describes the aim of this research project and the research questions, which this thesis sets out to address.

1.1 Social media in the context of eHealth

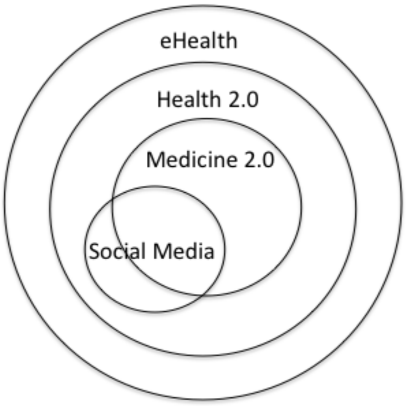
Social media-based health research forms part of the research field of eHealth. The term eHealth was initially introduced in 2000 and, although a variety of definitions exist, it can be characterised by the 'use of information technology in the delivery of health care' [1:pg 9]. Pagliari et al.'s scoping exercise to map the field of eHealth [2] indicates that most definitions emphasise the communicative functions of eHealth and specify the use of networked digital technologies. Neither study, however, goes so far as to specify which types of digital technologies are included in the definition. Since 2018, 'digital health' has become more commonly used following its inclusion as an agenda topic in the 71st World Health Assembly [3]. The WHO's *draft global strategy on digital health 2020-2024* [4] describes the use and scale-up of digital solutions to revolutionise how people worldwide achieve higher standards of health, and access services to promote and protect their health and well-being. The term is largely synonymous with eHealth.

Social media is defined as 'a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user generated content' [5: pg 61]. They include mainstream social media platforms for networking, crowdsourcing and collaboration, such as LinkedIn [6], Facebook [7], Twitter [8] and YouTube [9], as well as bespoke platforms aimed at specific health and patient communities such as PatientsLikeMe [10], TuDiabetes [11] and MedHelp [12]. These are sometimes referred to as 'e-health social networks' [13]. The terms

Health 2.0 and Medicine 2.0 have been coined to represent this convergence between Web 2.0 and healthcare.

In 2010, Van De Belt et al.’s systematic review identified 46 unique definitions of Health 2.0 and Medicine 2.0 and concluded that there is no general consensus on their scope, meaning and intention. They did, however, suggest that there were seven recurrent themes including Web 2.0, patients, professionals, social networking, health information, collaboration and change of healthcare and that Medicine 2.0 tends to focus on the relationship between the healthcare professional (HCP) and the patient. As such, Health 2.0 can be considered to be a sub-category of eHealth, and Medicine 2.0 a sub-category of Health 2.0 because of its focus solely on the HCP and patient relationship, to the exclusion of others. When considered in the context of eHealth, Health 2.0 and Medicine 2.0, social media can be categorised as a sub-category of Health 2.0, as illustrated in Figure 1.

Figure 1. Social media in the context of eHealth and Health 2.0



Pagliari et al.’s 2005 study indicated that eHealth has yet to be explicitly included amongst the taxonomic structure of medical and public health research databases such as Medline [2]. More than 10 years after this article was first published, the concepts of eHealth, Health 2.0 and Medicine 2.0 are still not represented in research databases and, instead, are distributed amongst a range of existing MeSH fields, including telemedicine and mobile health. The taxonomic structure provided by Medline introduced the hierarchical descriptor for social media in 2012 and describes them as ‘platforms that provide the ability and tools to create and publish information accessed via the Internet. Generally these platforms have three characteristics of user generated content, a high degree of interaction between creator and viewer, and can be easily integrated with other sites’ [14]. It classifies it, along with blogging, under the MeSH description of the Internet. The MeSH taxonomic hierarchy in which social media appears, however, is therefore considered relatively shallow and undifferentiated, even when compared to the taxonomic structures available in informatics research databases such as the ACM Library, where social media are classified under the category of World Wide Web. The next level of the hierarchy differentiates social media between ‘Web applications’, ‘Web interfaces’ and ‘Web searching and information discovery’ and is not specific to the application of these technologies in the context of health.

The term ‘social media’ itself is often used synonymously with ‘social networking’. In this thesis, social networking is considered to be one of eight social media available (Table 1) [5, 15].

Table 1. Types of social media

Type of Social Media	Description	Examples
Social networking sites	Individual users create personal information profiles and invite others to connect with them	Facebook
Blogs	Individual users create the equivalent of a personal web page that displays entries in reverse chronological order	WordPress
Microblogs	Individual users create short posts which are restricted by the number of characters or image uploads	Twitter
Wikis	Multiple users create joint and simultaneous content	Wikipedia
Social bookmarking	Multiple users create a group-based collection and rating of Internet content	Pinterest
Content communities	Individual users share media content such as photographs, videos, slides or audio files with other users	YouTube
Virtual worlds	Individual users appear in the form of an avatar and interact with others in a replicate of a three-dimensional environment	Second Life
Discussion forums	Multiple users create content on an online message board where people can hold conversations in the form of posted messages	Discussion forums on Macmillan.org.uk Diabetes.co.uk

1.2 Uses of social media in health research

Several uses of social media in health research have been described in the literature. These include the deployment of social media for the conduct of research, such as for gathering opinions [16], recruiting study participants [17], undertaking citizen science [18] or fostering stakeholder involvement [19], to name but a few. The benefits of its use have been identified as: increasing the diversity and number of people involved in research, accessibility, convenience and networking opportunities amongst contributors and minimal cost of use [20].

People's online activity on social media is also increasingly being used as a source of data for research [21]. Such secondary uses include studies seeking to profile or understand users' behaviours, demographics, interactions and networks or to assess their responses or sentiments towards particular topics, products or policies [22, 23]. One of the most significant trends, from both a scientific and a societal perspective, is the application of automated tools for mining and analysing social media as a means of revealing new associations or predicting future behaviours or outcomes. Increasingly, this is taking place alongside data mining from institutional or business repositories, to link historical and real-time information [24]. While the business sector has been using social media data for some time, such as to monitor brand reputation, their value for academic research is gradually being realised.

Bjerglund-Andersen and Söderqvist's typology [25] delineates these uses of social media in health research into five broad categories:

1. Research dissemination;
2. Scientific discussion and networking;
3. Engaging the public;
4. Academic teaching; and
5. Research and data collection.

The last of these can be divided into two qualitatively different categories – first, using social media platforms to *enable the conduct of research* and, secondly, using social media as a *source of data for research.*

1.3 Social media as a source of data for health research

Data created and shared on social media exist in text and visual content formats such as videos, photographs, pictures and emojis, which can be used for the purposes of research and health surveillance [26].

In 2004, the term 'infodemiology' was introduced by Eysenbach [27] to describe the science of distribution and determinants of information on the Internet, with the ultimate aim of informing public health and public policy. Data are collected and analysed in real-time and applications include monitoring status updates on social media. Eysenbach makes a distinction between supply- and demand-based infodemiology, where demand-based infodemiology is the search and navigation behaviour of people for online information. Supply-based infodemiology, on the other hand, is the continuous monitoring of user-generated content on the Internet and social media by employing natural language processing and other methods to classify posts by topic and obtain indicators of changes over time. The complementary term ‘infoveillance’ was introduced as a means of describing the automated and continuous analysis of unstructured, free text information available on the Internet and social media, for the purpose of public health surveillance. Despite very limited empirical evidence at the time, numerous potential applications were described, including early disease detection and surveillance of disease spread, as well as prevention and chronic disease management on a population level.

In Choi’s 2012 review of the past, present and future of public health surveillance [28], the author describes the components of public health surveillance as the development of a data framework, data collection, data analyses and interpretation and information dissemination. Although the future benefits of using the Internet in public health surveillance are discussed, no reference is made to social media as a potential source of this data. Despite this, infectious disease surveillance systems that extract and process social media data have become increasingly more common and include systems such as HealthMap, MedWatcher, Global Outbreak Alert and Response Network (GOARN), to name but a few.

With the increasing incidence and global burden of NCDs, the number of surveillance studies [29-33] that have been conducted to monitor and analyse the risk factors associated with these types of conditions have also increased. NCDs are defined by the WHO as those which do not pass from person to person, but are commonly linked to behavioural or physiological risk factors, such as tobacco use, physical inactivity, alcohol consumption and unhealthy diets [34]. NCDs are often referred to as chronic conditions whereby those who are diagnosed with them can, in some cases, live for a long time. They include cardiovascular diseases, cancers, chronic respiratory diseases and diabetes.

NCD surveillance studies often refer to the WHO’s STEPwise approach, which was introduced as part of its global surveillance strategy for preventing and controlling NCDs and their major risk factors. Data collected using standardised questions and protocols and information relevant to tobacco use, alcohol consumption, fruit and vegetable intake, obesity, blood pressure, cholesterol, exercise and diabetes is stored in a repository known as the NCD InfoBase, where all countries can use STEPS information for monitoring within-country and between-country trends [35]. The STEPwise approach, however, is limited by its use of the more traditional surveillance techniques of gathering data submitted via hospitals, clinics and local health authorities, before disseminating it for analysis. It does not take into account the large quantities of NCD-relevant data that are available on social media and could be used to compliment the health research and surveillance activities associated with these conditions.

In 2014, Velasco et al. published a systematic review into social media- and Internet-based data in global systems for infectious disease event-based surveillance [36]. The findings from this review suggest that the vast majority of these public health surveillance systems are classified as news aggregators and rely on media sources such as newspapers and broadcasters as their source of social media data. These systems are coordinated by organisations in cooperation with universities, NGOs and/or government agencies with the purpose of improving early detection, enhancing communication and collaboration and supplementing existing, more traditional surveillance systems. The findings from this review indicate that there are limitations to surveillance systems that extract data from social media and include the lack of content moderation before the data is disseminated for analysis and the ad-hoc frequency of data collection, as well as the lack of well-developed algorithms and statistical baselines. Velasco et al. were not able to identify a system that is currently part of a national program for surveillance and suggest that those that do exist are used intermittently as sources of information that are complementary to the more traditional public health surveillance systems.

Eggleston and Weitzman published a study in 2014 [37], which was a preliminary step towards addressing this gap in the literature. Although the study includes limited empirical evidence to validate their claims, the authors identify opportunities for preventative interventions for diabetes by augmenting information from traditional surveillance systems with data from social media with the distinction between passive and active social media-based surveillance. Passive social media-based surveillance is described as the mining of unstructured text with limited consideration for informed consent, transparency and validity, and active social media-based surveillance as a means of citizen science where patients voluntarily contribute health information to support disease discovery and treatment. Boman et al. [38] reiterate this, suggesting that research and surveillance that source data from social media can be used to complement more robust, but slower, methods of data collection and analysis.

1.4 Aim of this PhD research project

When this PhD research project was initiated in 2013, NCD research involving the secondary use of data from social media, whether it be for health research or surveillance purposes, was very limited. This thesis aims to contribute towards the growing body of eHealth knowledge and is intended to address the gap in empirical, evidence-based research that is associated with the secondary use of data from social media, to understand patient health issues and inform future health research into NCDs. As such, seven research questions were defined. Each of these are discussed in subsequent chapters of this thesis.

Research questions

RQ1: What is the status of the existing multi-disciplinary research literature based on analysis of data posted on social media for public health research and where are the gaps in this research? (Chapter 2)

RQ2: Can existing systematic review methods be re-purposed and applied to analyse data posted on social media platforms? (Chapter 3)

RQ3: How are research sponsors and researchers addressing the ethical challenges of analysing data posted on social media platforms? (Chapter 4)

RQ4: To what extent are diabetes-related posts on Twitter relevant to the clinical condition and what topics and intentions are represented in these posts? (Chapter 5)

RQ5: In what ways do people affected by Type 1 diabetes use different social media and what are the implications for health researchers? (Chapter 6)

RQ6: Are these differences in platform usage and associated data types also seen in people affected by lung cancer? (Chapter 7)

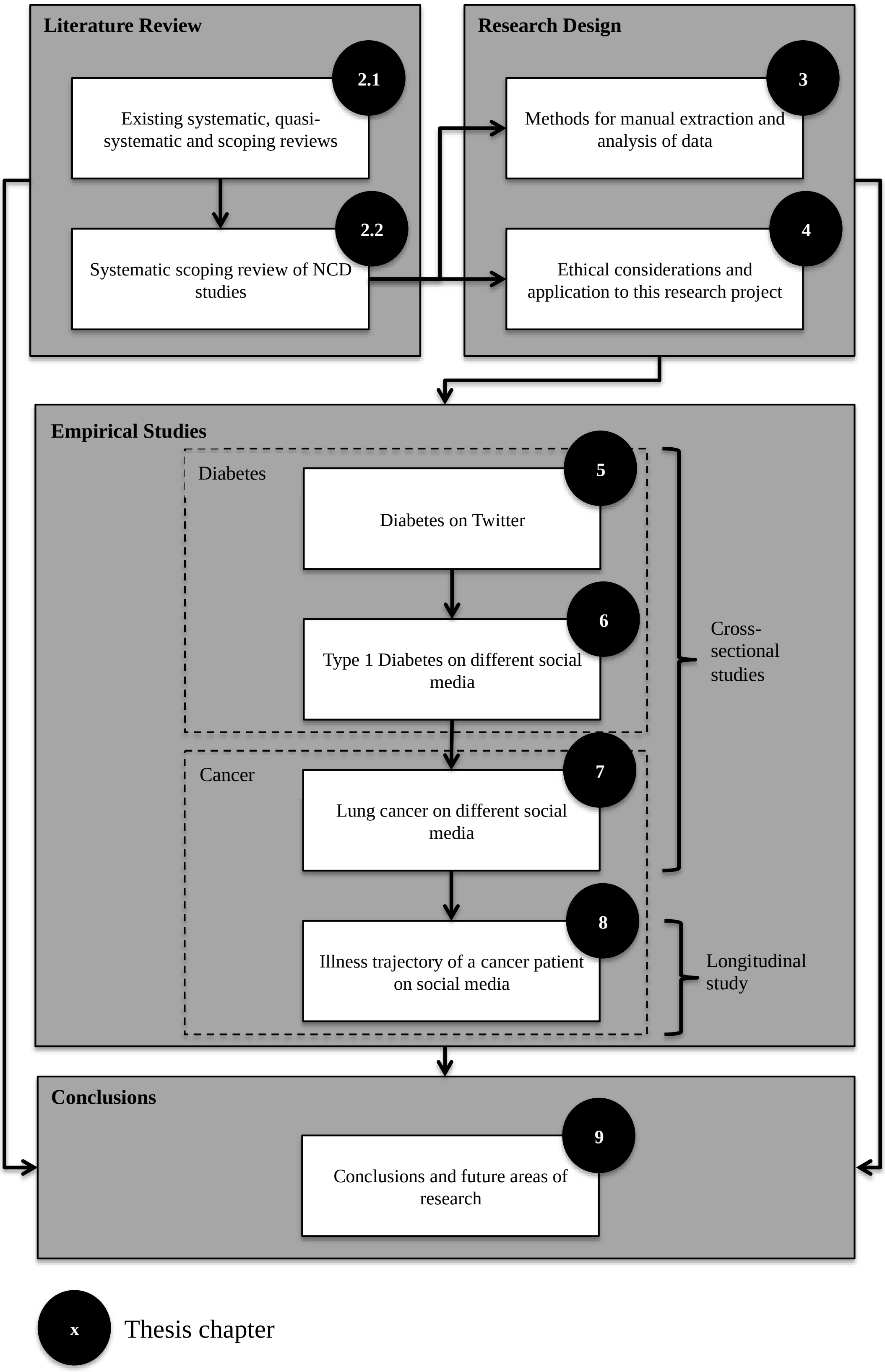
RQ7: Can characteristic illness trajectories be seen in a cancer patient’s digital narrative and what insights can be gained to inform palliative care services? (Chapter 8)

As described in Figure 2, RQ1 is discussed in the Literature Review (Chapter 2), which summarises the identified systematic, quasi-systematic and scoping reviews that investigate social media in the health context (Section 2.1) as well as the methods, results and conclusions of a systematic scoping review of NCD-related studies that source data from social media (Section 2.2). RQ2 (Chapter 3) focuses on re-purposing manual methods for the extraction and analysis of the data from social media.

RQ3 (Chapter 4) focuses on the ethical considerations of conducting studies that source data from social media, and includes a summary of the challenges and need for ethics as well as the methods, results, discussion, conclusions and implications from a study aimed at understanding how the RCUK's ethics guidelines address research involving social media. The findings of which were used to inform the research design of the diabetes and cancer-related empirical studies aimed at answering RQ4 (Chapter 5), RQ5 (Chapter 6), RQ6 (Chapter 7) and RQ7 (Chapter 8).

Each of these chapters describes the context and aim of the individual empirical study, the methods applied, the results, discussion, conclusions and implications for subsequent studies. Broader conclusions and areas for future research, as a consequence of this PhD research project, are discussed in Chapter 9.

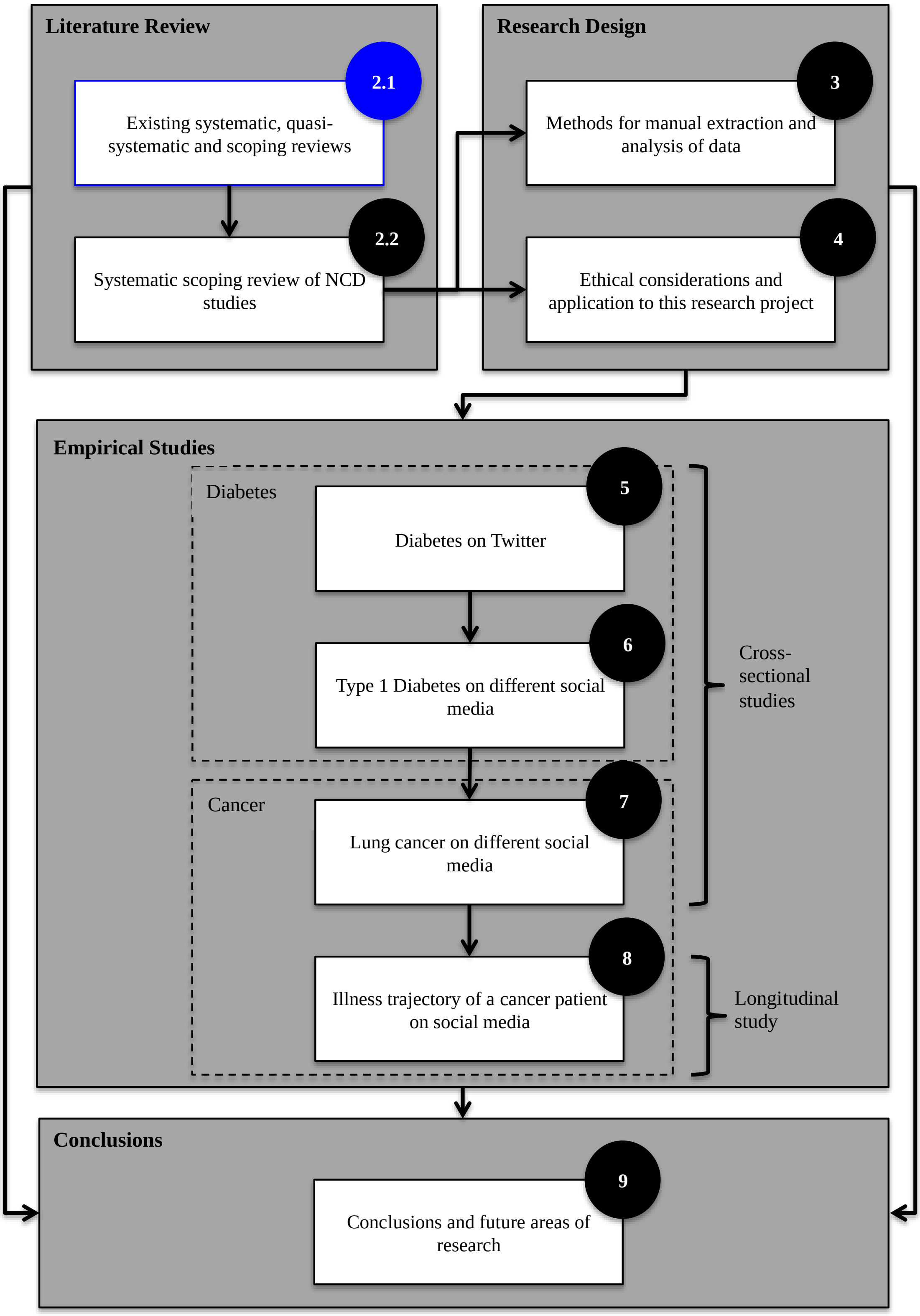
Figure 2. How to navigate this thesis



2. Literature review

The first known systematic review of the use of social media for public health practice and research was published in 2011 [39]. Since then, 49 systematic, quasi-systematic and scoping reviews have been identified, 24 of which are related to the secondary use of data from social media. Topics that were the subject of these reviews include infectious disease surveillance, NCDs/chronic disease, mental health, child health, vaccines, illicit drug use, adverse event monitoring, healthcare services, eGovernment, education, ethics and the use of social media as a communication and research tool. These reviews and their limitations are discussed in Section 2.1.

Despite the multi-disciplinary nature of this research field, the existing systematic and scoping reviews are limited by their narrow scope and limited consideration of the different health topics and disease, as well as the data, analytical methods and ethical issues involved. There has also been a tendency for research to be siloed within different academic communities (e.g. computer science, public health), hindering knowledge translation. In order to gain a deeper understanding of the landscape of existing literature, as well as to inform the empirical research for this thesis, a comprehensive systematic, scoping review was conducted with the aim of capturing and profiling a broad corpus of published multi-disciplinary research, in which data obtained from social media have been used to monitor, understand or evaluate aspects of health and disease. A sub-set of the literature, where NCDs are the subject of empirical investigation, were further classified and documented. The results of the scoping review, alongside the preliminary analyses undertaken to inform each stage, are detailed in Section 2.2 and are intended to address RQ1.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

2.1 Existing systematic, quasi-systematic and scoping reviews

The first known systematic, quasi-systematic or scoping review to investigate social media in the health context was published in 2011 with the intent to identify all the event-based surveillance systems that existed [39]. As of 2018, 49 such reviews have been identified following an iterative search of five electronic inter-disciplinary databases, including the Association of Computer Machinery (ACM), the Applied Social Sciences Index and Abstracts (ASSIA), the Institute of Electrical and Electronics Engineers (IEEE), PubMed and Scopus. Of these reviews, 24 focus on the secondary use of social media data for health research. The 49 reviews summarised in Table 2 are listed in chronological order. They include: approaches to infectious disease surveillance [39-48], NCDs and chronic disease [49-52], mental health [53-61], child health [62], vaccines [63], illicit drug use [64], adverse event monitoring [65], healthcare services [57, 66-68], patient and caregiver use [69], eGovernment [70], education [71], ethics [72-74] and the use of social media as a communication and research tool [75-85].

Table 2. Systematic, quasi-systematic and scoping reviews that investigate social media use in health

A= Infectious disease surveillance, B=Adverse event monitoring, C= Healthcare services, D=Illicit drug use, E=eGovernment, F=NCDs and chronic conditions including risk factors, G=Ethics, H=Mental health, I=Education, J=Vaccines, K= Communication and research tool, L= Child health, Y= Yes

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
1	Agheneza T. A Systematic Review of Event-Based Public Health Surveillance Systems. Hamburg: Faculty of Life Sciences, Hamburg University of Applied Sciences; 2011. [39]	To identify all the event-based surveillance systems that exist and where they are located.	3 (32)	• Thirteen event-based systems were identified and ten of the 13 systems evaluated. • North America is the leading continent with 77% of the event-based systems followed by Europe and lastly Asia. No system was found in Africa, Australia and South America.	• Focused solely on zoonotic diseases. • No protocol available. • Search conducted in 2011. • No bias assessment conducted.	Y	A
2	Guy S, Ratzki-Leewing A, Bahati R, Gwadry-Sridhar F. Social media: A systematic review to understand the evidence and application in infodemiology. Lect Notes Inst Comput Sci Soc Informatics Telecommun Eng. 2012; 91:1–8. [41]	To systematically review the literature utilising social media as a source for disease prediction and surveillance.	2 (12)	• Open-source micro-blogging sites can be used for influenza-like-illness monitoring. • Infoveillance can provide real-time, immediate and relevant information. • Pitfalls include classification of textual data, collected data may not be representative of the entire population and identification of geographical origin of data.	• Limited number of databases searched. • No protocol available. • Search conducted in 2011. • No meta-analysis or bias assessment conducted.	Y	A
3	Bernardo, T.M., Rajic, A., Young, I., Robiadek, K., Pham, M.T., Funk, J.A.	To assess the current state of knowledge regarding the use of search queries and social	1 database and 2 internet	• The use of search queries and social media for disease surveillance is a relatively recent	• Limited number of literature sources searched.	Y	A

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	Scoping review on search queries and social media for disease surveillance: a chronology of innovation J Med Internet Res. 2013; 15(7); e147. [42]	media for disease surveillance in order to inform future work on early detection and more effective mitigation of the effects of foodborne illness.	search engines (32)	phenomenon (first reported in 2006). • Tools and the methodologies for exploiting them are evolving over time. • Their accuracy, speed, and cost compare favourably with existing surveillance systems; the primary challenge is to refine the data signal by reducing surrounding noise.	• Focused solely on foodborne diseases. • Search conducted in 2012.		
4	Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res. 2013;15(2):e24. [66]	To identify frequently discussed issues about physician-rating websites.	7 (24)	• The current usage of physician-rating websites is still low but is increasing. • International data show that one out of six physicians has been rated, and approximately 90% of all ratings on physician-rating websites were positive. • Physicians should not ignore these websites, but rather, monitor the information available and use it for internal and external purposes.	• Did not investigate variations between different health topics or conditions. • No protocol available. • No bias assessment conducted. • Search conducted in 2012.	Y	C
5	Hamm MP, Chisholm A, Shulhan J, et al. Social media use among patients and caregivers: a scoping	To map the state of the existing literature evaluating the use of social media in patient and caregiver populations.	11 (284)	• There is an extensive and rapidly growing body of literature available. • Most studies have been descriptive; however, with such	• Did not investigate variations between different health topics or conditions.		C

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	review. BMJ Open 2013;3: e002819. [69]			widespread use, evaluations of effectiveness are needed. • In studies that have examined effectiveness, positive conclusions are often reported, despite the non-significant findings.	• Search conducted in 2012. • No protocol available.		
6	Jones. E, Sinclair J.M.A, R.I.G, and Barnard K.D. Social networking and understanding alcohol-associated risk for people with Type 1 diabetes: Friend or foe? Diabetes Technology & Therapeutics. 2013. [49]	To review the existing literature and to conduct a systematic search of the Internet to identify which social networking sites are being used by people with Type 1 diabetes.	4 databases 1 internet search engine (6)	• Young people with Type 1 diabetes are communicating via social media about their diabetes and are actively seeking advice on how to minimise the risks associated with alcohol consumption.	• Limited number (6) of studies identified. • No protocol available. • No meta-analysis or bias assessment conducted.	Y	F
7	Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. Journal of medical Internet research, 15(4), e85. [75]	To review the published literature to identify the uses, benefits, and limitations of social media for health communication among the general public, patients, and health professionals.	9 (98)	• Seven main uses of social media identified as: (1) provide health information on a range of conditions, (2) provide answers to medical questions, (3) facilitate dialogue between patients to patients, and patients and health professionals, (4) collect data on patient experiences and opinions, (5) provide health intervention, health promotion and health education, (6) reduce stigma and (7) provide online consultations.	• No protocol available and no explanation how themes used in the synthesis were identified. • Did not investigate variations between different health topics or conditions.		K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				• Six benefits were identified as (1) increased interactions with others, (2) more available, shared and tailored information, (3) increased accessibility and widening access to health information, (4) peer/social/emotional support, (5) public health surveillance and (6) potential to influence health policy. • Twelve limitations were identified, primarily consisting of quality concerns and lack of reliability, confidentiality and privacy.			
8	Park, B. K., & Calamaro, C. (2013). A systematic review of social networking sites: Innovative platforms for health research targeting adolescents and young adults. Journal of Nursing Scholarship, 45(3), 256-264. [53]	To review evidence to determine if social networking sites are effective tools for health research in the adolescent and young adult populations.	7 (17)	• Advantages of using social networking sites apparent in this review are (1) ease of access to youth, (2) cost-effectiveness in recruitment, (3) ease of intervention and (4) reliable screening venue of mental status and high-risk behaviors.	• Did not investigate variations between different health topics or conditions. • No protocol available.	Y	H
9	Best, P., Manktelow, R., & Taylor, B. (2014). Online communication, social media and adolescent wellbeing: A systematic narrative	To systematically review and synthesise current empirical research, identifying both the beneficial and harmful effects of online	8 (43)	• The benefits of using online technologies were reported as increased self-esteem, perceived social support, increased social capital, safe identity	• Did not investigate variations between different health topics or conditions.		H

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	review. Children and Youth Services Review, 41, 27-36. [54]	communication and social media technology amongst young people.		experimentation and increased opportunity for self-disclosure. • Harmful effects were reported as increased exposure to harm, social isolation, depression and cyber-bullying. • The majority of studies reported either mixed or no effect(s) of online social technologies on adolescent wellbeing.	• No protocol available.		
10	Capurro, D., Cole, K., Echavarria, M., Joe, J., Neogi, T., & Turner, A. (2014). The Use of Social Networking Sites for Public Health Practice and Research: A Systematic Review. Journal Of Medical Internet Research, 16(3), 213-226. [76]	To conduct a systematic literature review to identify the use of social networking sites for public health research and practice and to identify existing knowledge gaps.	3 (73)	• The number of publications about public health uses for social networking sites has been steadily increasing in the past five years. • With few exceptions, the literature largely consists of observational studies describing users and usages of social networking sites regarding topics of public health interest. • A large proportion of the identified studies included populations considered hard to reach, such as young individuals, adolescents and individuals at risk of sexually transmitted diseases or alcohol and substance abuse.	• Search strategy for this review did not include discussion forums, blogs or health specific social media sites. • No protocol available. • No assessment of bias conducted.		K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
11	Conway, M. (2014). Ethical issues in using Twitter for public health surveillance and research: Developing a taxonomy of ethical concepts from the research literature. Journal of Medical Internet Research, 16(12), E290. [73]	To develop a taxonomy of public health surveillance-related ethical concepts that emerge when using Twitter data, with a view to: (1) explicitly identifying a set of potential ethical issues and concerns that may arise when researchers work with Twitter data and (2) providing a starting point for the formation of a set of best practices for public health surveillance through the development of an empirically derived taxonomy of ethical concepts.	4 (13)	• Iteratively generated a taxonomy of ethical concepts consisting of ten top-level categories: privacy, informed consent, ethical theory, institutional review board (IRB)/regulation, traditional research vs Twitter research, geographical information, researcher lurking, economic value of personal information, medical exceptionalism and benefit of identifying socially harmful medical conditions.	• No protocol available. • No assessment of bias conducted. • Did not investigate variations between different health topics or conditions.		G
12	Cotter, A., Durant, N., Agne, A., & Cherrington, A. (2014). Internet interventions to support lifestyle modification for diabetes management: A systematic review of the evidence. Journal of Diabetes and Its Complications, 28(2), 243-251. [50]	To identify studies that used Internet-based interventions to promote diabetes education and lifestyle modification among adults with Type 2 diabetes. In light of existing diabetes related health disparities that exist along socio-economic and racial/ethnic lines, a secondary objective was to	1 (9)	• Two studies demonstrated improvements in diet and/or physical activity and two studies demonstrated improvements in glycemic control comparing web-based intervention with control. • Successful studies were theory-based, included interactive components with tracking and personalized feedback, and	• Limited number of databases searched. • Limited number of studies (nine) identified. • No protocol available.		F

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
		assess the extent to which these interventions are tailored for diverse and/or underserved communities.		provided opportunities for peer support. • Few studies focused on high risk, underserved populations.			
13	Gustafson, D., & Woodworth, C. (2014). Methodological and ethical issues in research using social media: A metamethod of Human Papillomavirus vaccine studies. BMC Medical Research Methodology, 14(1), 127. [74]	To analyse primary research studies that used social media to explore knowledge about Human Papillomavirus infections and vaccination and to clarify, synthesise and reflect on the possibilities and ethical considerations of using social media as a data collection tool in health research.	7 (6)	• Designs of the six selected studies were sound, although most studies could have been more transparent about how they built in rigor to ensure the trustworthiness and credibility of findings. • Statistical analysis that intended to measure trends and patterns did so without the benefit of randomised sampling and other design elements for ensuring generalisability or reproducibility of findings beyond the specified virtual community. • Most researchers did not sufficiently engage virtual users in the research process or consider the risk of privacy incursion. • Most studies did not seek ethical approval from an institutional research board or permission from host websites or web service providers.	• Limited number of studies (six) identified. • No protocol available. • No assessment of bias conducted. • Did not investigate variations between different health topics or conditions.		G

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
14	Hamm, M. P., Shulhan, J. D., Williams, G., Milne, A., Scott, S., & Hartling, L. (2014). A systematic review of the use and effectiveness of social media in child health. Acta Veterinaria Scandinavica, 14(1), 138.[62]	To conduct a systematic review to determine: 1) for what purposes social media is being used in child health and its effectiveness; and 2) the attributes of social media tools that may explain how they are or are not effective.	11 (25)	• The most common uses for social media were for health promotion with a focus on healthy diet and exercise, sexual health, smoking cessation and parenting issues. • Adolescents were the most common target audience, discussion forums were the most commonly used tools and the tools were largely community-based. • Nearly all studies concluded that the social media tool evaluated showed evidence of utility; however, results of the primary outcomes from the majority of comparative studies showed no significant effect.	• Search conducted in 2012. • No protocol available.		L
15	Laaksonen, C., Jalonen, H., & Paavola, J. (2014). Utilising Social Media for Intervening and Predicting Future Health in Societies. Communications in Computer and Information Science, 450, 100-108.[77]	To identify 1) systematic reviews describing the relation between social media and health and 2) previous research on utilising social media for predicting health on a population level.	1 (4)	• No systematic review on utilising social media to predict health on population levels were found and the area of research is new. • The included articles on utilising social media to predict health in populations were published between 2009–2013. The articles described identifying early warning signs with regard to health status in populations,	• Search conducted in 2013. • Limited number of studies (four). • Only one data source included. • No protocol available. • No bias assessment conducted.	Y	K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				predicting national suicide incidents, vaccination control and introducing a new framework of infoepidemiology to predict relevant health events in populations. • A common conclusion of the reviews was that social media and new technology may carry potential for health outcomes, but more research is needed to understand the relation between social media and health related issues as well as to evaluate the effects, especially the long-term outcomes.			

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
16	Smith, T., & Lambert, R. (2014). A systematic review investigating the use of Twitter and Facebook in university-based healthcare education. Health Education, 114(5), 347-366. [71]	To evaluate the use, attitudes and perceptions of both teachers and students towards social media platforms (Facebook and Twitter) in healthcare higher education practice.	13 (16)	• Facebook and Twitter are perceived by students and lecturers as useful adjuncts in healthcare higher education practice. This review found that Facebook and Twitter when used, as part of a 'blended' approach to classroom teaching, has been demonstrated to enhance communication and increase accessibility to real-world practices and expertise. • Whilst students perceive this to be of value, the role of faculty members in a predominantly 'social' community has been acknowledged as a potential conflict. Similarly, lecturers have also reported concerns regarding the perceived boundaries between student-staff relationships. Both students and staff have concerns regarding e-professionalism and the potential for adverse exposure to colleagues, employers and patient groups.	• Limited to two social media platforms. • No protocol available. • Did not investigate variations between different health topics or conditions.		I

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
17	Velasco, E., Agheneza, T., Denecke, K., Kirchner, G., Eckmanns, T., Social media and internet-based data in global systems for public health surveillance: a systematic review. Milbank Q. 2014; 92(1): 7-33 [43]	To explore how useful data from social media and the Internet are, and what the potential is to enhance surveillance.	3 (13)	• No event-based surveillance systems are currently used in national surveillance programs. • Existing systems have limitations. • There is a need for further development of automated technologies that monitor health-related information on the Internet. • The dissemination to health authorities of new information about health events is not always efficient and could be improved. • No comprehensive evaluations show whether event-based surveillance systems have been integrated into actual epidemiological work during real-time health events.	• Did not consider other health topics or conditions (only infectious disease). • Did not consider geographical scope. • Search conducted in 2013.	Y	A
18	Charles-Smith, L.E., Reynolds, T., Cameron, M., Conway, M., Lau, E.H.Y., Olsen, J.M., Pavline, J.A., Shigematsu, M., Streichert, L.C., Suda, K.J., Corley, C.D., (2015). Using Social Media for Actionable Disease Surveillance and Outbreak Management: A	To determine whether social media can be integrated into disease surveillance practice and outbreak management to support and improve public health and if it can be used to effectively target populations, specifically	4 (60)	• Topics covered in the studies include: infectious disease, non-infectious disease, medication and vaccines, health risk behaviours. • Identified many evidence gaps and biases in the current knowledge.	• Search conducted in 2013.	Y	A

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	Systematic Literature Review. 10(10) [44]	vulnerable populations, to test an intervention and interact with a community to improve health outcomes.		• There are few studies to date on interventions and a lack of use of social media in practice despite the high potential for success identified in exploratory studies. • Findings suggest that it is particularly challenging to translate research using social media for biosurveillance into practice. The focus of many studies, especially on infectious diseases, is done retrospectively, potentially highlighting the ease in prediction post outbreak rather than implementation of social media prospectively.			
19	Golder, S., Norman, G., Loke, Y.K. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media, British Journal of Clinical Pharmacology. 2015; 80(4); 878-88 [65]	To summarise the prevalence, frequency and comparative value of information on the adverse events of healthcare interventions from user comments and videos in social media.	16 databases and 2 internet search engines (51)	• The prevalence of adverse events reports on social media varies from 0.2% to 8% of posts. • 'Mild' and symptom-related adverse events are over-represented in social media and laboratory test abnormalities and 'serious' adverse events are under-represented compared with other data sources. • The question as to whether searching social media for adverse events data is a valuable	• No protocol available. • Did not focus on any specific condition, health topic, country or social media platform. • Retrospective studies.	Y	B

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				use of resources, resulting in improved patient outcomes, remains unanswered.			
20	Hu, Y. (2015). Health communication research in the digital age: A systematic review. Journal of Communication in Healthcare, 8(4), 260-288. [78]	To examine patterns and trends of research pertaining to the intersection between technology and health communication, to determine where gaps exist within the current body of literature and to provide future directions for researchers and practitioners.	3 (348)	• The most investigated subject in the literature was online health information, including search, dissemination and evaluation, followed by telemedicine/electronic medical records, social support, the provider–patient relationship and health intervention. • The most studied health concerns were cancer and mental health, followed by HIV/AIDS/STD, general nutrition, diet, exercise/obesity, physical disability, diabetes and substance/alcohol/tobacco use. • Two out of 348 studies looked at social media.	• Limited focus on social media, limiting extrapolation of findings. • No protocol available.		K
21	Odone, A., Ferrari, A., Spagnoli, F., Visciarelli, S., Shefer, A., Pasquarella, C., & Signorelli, C. (2015). Effectiveness of interventions that apply new media to improve vaccine uptake and vaccine coverage: A systematic review. Human	To systematically collect and summarise the available evidence on the effectiveness of interventions that apply new media to promote vaccination uptake and increase vaccination coverage.	2 (19)	• The majority of the studies were conducted in the US (74%). • Retrieved studies explored the role of text messaging (37%), smart- phone applications (5%), YouTube videos (5%), Facebook (5%), targeted websites and portals (21%), software for	• Limited focus on social media, limiting extrapolation of findings. • No protocol available.		J

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	Vaccines & Immunotherapeutics, 11(1), 72-82. [63]			physicians and health professionals (21%) and email communication (5%). • There is some evidence that text messaging, accessing immunisation campaign websites, using patient-held web-based portals and computerised reminders increase immunisation coverage rates. • Insufficient evidence is available on the use of social networks, email communication and smartphone applications.			
22	Patel, Chang, Greysen, & Chopra. (2015). Social Media Use in Chronic Disease: A Systematic Review and Novel Taxonomy. The American Journal of Medicine, 128(12), 1335-1350. [86]	To evaluate clinical outcomes from applications of contemporary social media in chronic disease; to develop a conceptual taxonomy to categorize, summarise and then analyse the current evidence base; and to suggest a framework for future studies	1 (42)	• The overall impact of social media on chronic disease was variable. • Among studies that showed benefit, most used either Facebook or blogs, and were based within the domain of support.	• No protocol available. • Limited number of databases.	Y	F
23	Al-Garadi, Khan, Varathan, Mujtaba, & Al-Kabsi. (2016). Using online social networks to track a pandemic: A systematic	To conduct a systematic review of the use of online social networks to track a pandemic.	5 (20)	• Identified the extent of using online social networks to track a pandemic.	• Limited conditions investigated. • No protocol available.	Y	A

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	review. Journal of Biomedical Informatics, 62, 1-11. [45]			• Online social networks contain significant information to track a pandemic almost in real time. • There are challenges and future implications.			
24	Baker, D., & Algorta, G. (2016). The Relationship Between Online Social Networking and Depression: A Systematic Review of Quantitative Studies. Cyberpsychology, Behavior, and Social Networking, 19(11), 638-648. [55]	To examine current quantitative studies focused on the relationship between online social networking and symptoms of depression.	5 (30)	• The relationship between online social networking and symptoms of depression may be complex and associated with multiple psychological, social, behavioural and individual factors. • The impact of online social networking on wellbeing may be both positive and negative, highlighting the need for future research.	• Did not investigate variations between different health topics or conditions. • No protocol available.		H
25	Carvalho, R., & Teixeira Lopes, C. (2016). Social network analysis to understand behaviour dynamics in online health communities: A systematic review. 2016 11th Iberian Conference on Information Systems and Technologies (CISTI), 2016, 1-7. [85]	To better understand the most popular approaches and methods for analysing behaviour dynamics of online health communities using social network analysis	3 (13)	• There are few studies that make use of the Social Network Analysis (i.e. graph detection) as a singular technique to analyze an online health community. It is typically combined with additional metrics and/or tools that hold higher-level analytics such as content analysis tools. • A common methodology among researchers conducting Social Network Analysis on online	• Not a structured systematic review. • No protocol available. • Limited conditions and health topics. • No assessment of bias.	Y	K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				health communities is to analyse the network in terms of density, weight of arcs and nodes and apply typical centrality measures. • The number of studies regarding Social Network Analysis to communities in the health domain was significantly higher when addressing physical – non-virtual – communities.			
26	Fung, Tse, Cheung, Miu, & Fu. (2014). Ebola and the social media. The Lancet, 384(9961), 2207. [87]	To systematically review existing research pertinent to Ebola virus disease and social media, especially to identify the research questions and the methods used to collect and analyse social media.	6 (13)	• Included seven articles on Twitter, one on Facebook, three on YouTube and one on Instagram & Flickr. • Elevent articles examined Ebola posts' content themes, meta-data or user characteristics. • One article examined how Ebola-related news videos influenced Twitter traffic.	• No structured methodology applied. • Limited number of studies (11) identified.	Y	A
27	Gabarron, E., & Wynn, R. (2016). Use of social media for sexual health promotion: A scoping review. Global Health Action, 9(1), 32193. [79]	To review the scientific literature on the use of online social media for sexual health promotion.	10 (51)	• Fifty-one publications on the use of social media for promoting sexual health identified. • About a quarter of the publications have identified promising results and the evidence for positive effects of social media interventions for	• No protocol available. • Limited conditions and health topics. • No assessment of bias.		K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				promoting sexual health is increasing.			
28	Robinson, J., Cox, G., Bailey, E., Hetrick, S., Rodrigues, M., Fisher, S., & Herrman, H. (2016). Social media and suicide prevention: A systematic review. Early Intervention in Psychiatry,10(2), 103-121. [56]	To identify current evidence pertaining to the ways in which social media are currently used as a tool for suicide prevention.	5 (30)	• Thirty studies were included; four described the development of social media sites designed for suicide prevention, six examined the potential of social media in terms of its ability to reach or identify people at risk of suicide, 15 examined the ways in which people used social media for suicide prevention-related purposes and five examined the experiences of people who had used social media sites for suicide prevention purposes.	• No protocol available. • Limited conditions and health topics investigated. • No assessment of bias.		H
29	Smailhodzic, E., Hooijsma, W., Boonstra, A., & Langley, D. (2016). Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. Bmc Health Services Research, 16(1), 442. [67]	To provide an overview of literature on the effects of social media use for health-related reasons on patients and their relationship with healthcare professionals.	2 (23)	• Six categories of patients' use of social media were identified, namely: emotional, information, esteem, network support, social comparison and emotional expression. • The types of use were found to lead to seven identified types of effects on patients, namely improved self-management and	• No protocol available. • Did not investigate variations between different health topics or conditions. • No assessment of bias.	Y	C

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				control, enhanced psychological well-being, and enhanced subjective well-being, diminished subjective well-being, addiction to social media, loss of privacy and being targeted for promotion. • Social media use by patients was found to affect the healthcare professional and patient relationship by leading to more equal communication between the patient and healthcare professional, increased switching of doctors, harmonious relationships and suboptimal interaction between the patient and healthcare professional. • Five out of 22 studies analysed data from social media platforms			
30	Välimäki, M., Athanasopoulou, C., Lahti, M., & Adams, C. (2016). Effectiveness of Social Media Interventions for People With Schizophrenia: A Systematic Review and Meta-Analysisa. Journal of Medical Internet Research, 18(4), E92. [57]	To determine the effects of social media interventions for supporting mental health and well-being among people with schizophrenia.	10 (2)	• Only two studies identified out of 1,043, whereas numerous reports have been published citing the benefits of social media in mental health. • Findings suggest the effects of social media interventions are largely unknown.	• Limited studies identified (two). • No protocol available.		H

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				• More research is needed to understand the effects of social media, for users with and without mental illness, in order to determine the impact on mental well-being of social media use as well as its risks.			
31	Verhoef, L.M., Van de Belt, T.H., Engelen, L.J.L.P.G., Schoonhoven, L., Kool, R.B., Social media and rating sites as tools to understanding quality of care: a scoping review, J Med Internet Res. 2014; 16(2); e56 [57, 68]	To explore the relationship between data from social media and quality of care.	4 (29)	• There is a relationship between information on social media and quality of health care • Drawbacks exist, since the rating is anonymous and vulnerable to fraud. They are often based on only a few reviews and are predominantly positive. • People providing feedback on health care via social media are presumably not always representative of the patient population.	• No protocol available. • No specific conditions.	Y	C
32	Zhou, Xia, Yap, Peiling, Tanner, Marcel, Bergquist, Robert, Utzinger, Jürg, & Zhou, Xiao-Nong. (2016). Surveillance and response systems for elimination of tropical diseases: Summary of a thematic series in Infectious Diseases of Poverty. Infectious Diseases of Poverty, 5(1), 49. [46]	To review a thematic series on surveillance-response systems for elimination of tropical diseases.	1 Journal (13)	• Twenty-two contributions covering a broad array of diseases are featured – i.e. clonorchiasis, dengue, hepatitis, human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS), H7N9 avian influenza, lymphatic filariasis,	• Not a structured systematic review. • No protocol available.		A

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				malaria, Middle East respiratory syndrome (MERS), rabies, schistosomiasis and tuberculosis (TB). • There are five scoping reviews, a commentary, a letter to the editor, an opinion piece and an editorial pertaining to the theme 'Elimination of tropical disease through surveillance and response'. • The remaining 13 articles are original contributions mainly covering (i) drug resistance; (ii) innovation and validation in the field of mathematical modelling; (iii) elimination of infectious diseases; and (iv) social media reports on disease outbreak notifications released by national health authorities. • Analysis of the authors' affiliations reveals that scientists from the People's Republic of China are prominently represented.			
33	Calvo, R., Milne, D., Hussain, M., & Christensen, H. (2017). Natural language processing in mental	To highlight areas of research where NLP has been applied in the mental	1 search engine	• Most popular forms of social media have been used as data	• Limited number of data sources.	Y	H

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	health applications using non-clinical texts †. 23(5), 649-685. [58]	health literature and to help develop a common language that draws together the fields of mental health, human-computer interaction and NLP.	(23)	sources for mental health applications. • Number of users, language (i.e. English) and availability of APIs increase the chances of a platform being used. Twitter is the most widely used source of data mainly because the collection of public data is easy. Facebook is also common, often used by authors who also work for (or in partnership with) the company. • Only searched papers in English; these mostly talked about content written in English. A few exceptions (e.g. Japanese) are mentioned. It would seem that NLP in non-English languages is an unexplored area. This may be related to the lower quality or absence of NLP tools in languages other than English. • All feature selection and classification algorithms common in the NLP literature have been tried.	• Did not investigate variations between different health topics or conditions. • No protocol available. • No bias assessment.		

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
34	Carson, Kv, Ameer, F, Sayehmiri, K, Hnin, K, Van Agteren, Jem, Sayehmiri, F, Brinn, M.P., Esterman, A.J, Smith, Bj. (2017). Mass media interventions for preventing smoking in young people. Cochrane Database Of Systematic Reviews, 6(6) [80]	To assess the effects of mass media interventions on preventing smoking in young people, and whether it can reduce smoking uptake among youth (under 25 years), improve smoking attitudes, intentions and knowledge, improve self-efficacy/self-esteem, and improve perceptions about smoking, including the choice to follow positive role models.	3 (8)	• Three out of eight studies found that the intervention was effective in preventing smoking in youth. The remaining five studies did not detect an effect • Seven studies were conducted in the USA and one was conducted in Norway. • One interrupted time-series study of a social media intervention.	• Limited number of studies (eight) identified.		K
35	Coto, M., Lizano, F., Mora, S., & Fuentes, J. (2017). Social media and elderly people: Research trends. Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10283, 65-81 [88]	A systematic review of studies on social media and older adults to understand the kind of research approaches that are been used with this population, the data collection methods, which are the more prevalent research domains, the main research goals and obtained results.	10 databases 1 search engine (36)	• The study included 36 research papers. • There is currently no predominant research approach to address this field of study. Samples were generally small and questionnaires were the most common method of data collection. Many of the research efforts are focused on aspects of design, the interaction of the elderly with social media and how it affects their lives.	• Did not investigate variations between different health topics or conditions. • No protocol available. • No assessment of bias.		K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				• Many papers do not define social media and an important number do not indicate the research approach nor the number of participants.			
36	Daher, J., Vijh, R., Linthwaite, B., Dave, S., Kim, J., Dheda, K., Peter, T., Pai, N.P. (2017). Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). Bmj Open, 7(11) [81]	To evaluate the feasibility and impact of all digital innovations for all HIV/STIs.	4 (99)	• Internet-enabled mHealth and other internet-based eHealth innovations were evaluated in 21% (21/99) of studies. These innovations consisted of many different forms: social media and online campaigns (9/21), avatar-guided computer programs (2/21), mobile applications (5/21), combination of social media and websites (2/21), websites (1/21), telemedicine services (1/21) and streamed soap opera videos (1/21). • By geographical location, 37% (37/99) of studies were conducted in North America, 26% (26/99) in Sub-Saharan Africa, 24% (24/99) in Europe, 7% (7/99) in Oceania, 3% (3/99) in South-East Asia and 2% (2/99) in South America. • Innovations need to be proven for impact and cost-effectiveness,	• Did not investigate variations between different health topics or conditions. • No protocol available.		K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				using a combination of clinical trials, quasi-randomised studies, observational studies and qualitative research studies.			
37	Golder S, Ahmed S, Norman G, Booth A., Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review, J Med Internet Res 2017;19(6):e195 [72]	To ascertain attitudes towards the ethical considerations of using social media as a data source for research as expressed by social media users and researchers.	16 data bases and 2 internet search engines (17)	• Attitudes varied from overly positive, with people expressing the views about the essential nature of such research for the public good, to very concerned with views that social media research should not happen. • Underlying reasons for this variation related to issues such as the purpose and quality of the research, researcher affiliation and potential harms. • The methods used to conduct the research were also important. • Many respondents were positive about social media research, while adding caveats such as the need for informed consent or use restricted to public platforms only.	• Limited conditions and health topics.	Y	G
38	Kazemi, D., Borsari, B., Levine, M., & Dooley, B. (2017). Systematic review of surveillance by social media platforms for illicit drug use.	To review the ability of social media to better recognise illicit drug use trends.	9 (14)	• Selected studies used a range of social media tools/applications, including message boards, Twitter and blog/forums/platform discussions.	• No protocol available. • No bias assessment.	Y	D

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	Journal of Public Health (Oxford, England), 1-14. [64]			• Limitations included relevance, a lack of standardised surveillance systems and a lack of efficient algorithms to isolate relevant items. • Systematic approaches need to be developed to efficiently extract and analyse illicit drug content from social networks to supplement effective prevention programs.			
39	Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K., John, A. (2017). A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS ONE, 12(8), E0181722. [59]	To systematically review evidence regarding the potential influence of the internet on self-harm/suicidal behaviour in young people.	10 (46)	• There is significant potential for harm from online behaviour (normalisation, triggering, competition, contagion) but also the potential to exploit its benefits (crisis support, reduction of social isolation, delivery of therapy, outreach). • Young people appear to be increasingly using social media to communicate distress, particularly to peers. • The focus should now be on how specific mediums' (social media, video/image sharing) might be used in therapy and recovery.	• Limited number of social media studies (4/46). • Did not investigate variations between different health topics or conditions.		H

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
40	Nour, M., Yeung, S., Partridge, S., & Allman-Farinelli, M. (2017). A Narrative Review of Social Media and Game-Based Nutrition Interventions Targeted at Young Adults. Journal of the Academy of Nutrition and Dietetics, 117(5), 735-752.e10. [51]	To evaluate the efficacy of social media and game-based interventions in nutrition promotion and behaviour modification.	10 (11)	• Social media and gaming offer a new dimension for nutrition interventions, with the current body of evidence indicating potential positive impacts on improving knowledge and attitudes. • The implications of social media and gaming strategies in the longer-term and for influencing behaviour and health outcomes could not be determined.	• Limited conditions, health topics and population segment.		F (risk factor)
41	O'Shea, Jesse. (2017). Digital disease detection: A systematic review of event-based internet biosurveillance systems. International Journal of Medical Informatics, 101, 15-22. [47]	To update the current state of knowledge on event-based Internet biosurveillance systems by identifying all systems, including current functionality, with hopes to aid decision makers with whether to incorporate new methods into comprehensive programmes of surveillance.	3 (99)	• Fifty event-based Internet systems identified. • Each system uses different innovative technology and data sources to gather data, process and disseminate data to detect infectious disease outbreaks.	• No specific condition investigated. • No protocol available. • No assessment of bias.	Y	A

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
42	Overbey, K., Jaykus, L., & Chapman, B. (2017). A Systematic Review of the Use of Social Media for Food Safety Risk Communication. Journal Of Food Protection, 80(9), 1537-1549. [82]	To analyse literature recommendations and draw conclusions about how best to utilise social media for food safety risk communication going forward.	3 (24)	• Trust and personal beliefs were important drivers of social media use. • The wide reach, immediacy and information-gathering capacities of social media were frequently cited benefits. Suggestions for social media best practices were inconsistent among studies and study designs were highly variable. • More evidence-based suggestions are needed to better establish guidelines for social media use in food safety and infectious disease risk communication. The information gleaned from this review can be used to create effective messages for shaping food safety behaviours.	• No protocol available. • Limited number of studies - Of the 24 articles, 16 dealt with food safety communications; the other 8 articles dealt with communications related to infectious diseases.		K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
43	Rose, T., Barker, M., Maria Jacob, C., Morrison, L., Lawrence, W., Strömmer, S., Vogel, C., Woods-Townsend, K., Farrell, D., Inskip, H., Baird, J. (2017). A Systematic Review of Digital Interventions for Improving the Diet and Physical Activity Behaviors of Adolescents. Journal of Adolescent Health, 61(6), 669-677.[52]	To synthesise evidence on the effectiveness of digital interventions to improve diet quality and increase physical activity in adolescents, to effective intervention components and to assess the cost-effectiveness of these interventions.	8 (27)	• Most (n=15) were website interventions. Other delivery methods were text messages, games, multicomponent interventions, emails, and social media. • Significant behaviour change was often seen when interventions included education, goal setting, self-monitoring and parental involvement. • None of the publications reported cost-effectiveness. Due to the heterogeneity of the studies, meta-analysis was not feasible. • It is possible to effect significant health behaviour change in adolescents through digital interventions that incorporate education, goal setting, self-monitoring and parental involvement.	• Limited number of studies included - one study in 26 related to social media. • No protocol available • No assessment of bias conducted.		F (risk factor)
44	Sinnenberg, L., Buttenheim, A., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. (2017). Twitter as a Tool for Health Research: A Systematic Review. American Journal Of	To systematically review the use of Twitter in health research, define a taxonomy to describe Twitter use and characterise the current state of Twitter in health research.	5 (137)	• Of 1,110 unique health-related articles mentioning Twitter, 137 met the eligibility criteria. • The primary approaches for using Twitter in health research that constitute a new taxonomy were content analysis (56%; n =	• Did not investigate variations between different health topics or conditions. • No protocol available.	Y	K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
	Public Health, 107(1), E1-E8.[83]			77), surveillance (26%; n = 36), engagement (14%; n = 19), recruitment (7%; n = 9), intervention (7%; n = 9) and network analysis (4%; n = 5). • These studies collectively analysed more than 5 billion tweets, primarily by using the Twitter application program interface.	• No assessment of bias conducted.		
45	Tursunbayeva, Franco, & Pagliari. (2017). Use of social media for e-Government in the public health sector: A systematic review of published studies. Government Information Quarterly, 34(2), 270-282, [70]	To investigate the adoption and use of social media by public health organisations.	4 (22)	• Evidence-base is mostly descriptive, unidisciplinary and lacks the theoretical depth seen in other branches of e-Government research. • The lack of robust evidence makes it difficult to draw conclusions about the effectiveness of these approaches in the public health sector.	• Did not investigate variations between different health topics or conditions.	Y	E
46	Twomey, C., & O'Reilly, G. (2017). Associations of Self-Presentation on Facebook with Mental Health and Personality Variables: A Systematic Review. Cyberpsychology, Behavior, and Social Networking, 20(10), 587-595. [60]	To (1) identify mental health and personality variables associated with self-presentation on Facebook; (2) interpret arising trends in self-presentation associations; (3) investigate the operationalisation of self-	4 (21)	• Significant self-presentation associations were yielded for self-esteem, perceived social support, social anxiety, well-being, depression, bipolar/mania, stress, self-consciousness and insecure attachment from 21 studies.	• One social media platform included. • No protocol available. • No assessment of bias conducted.		H

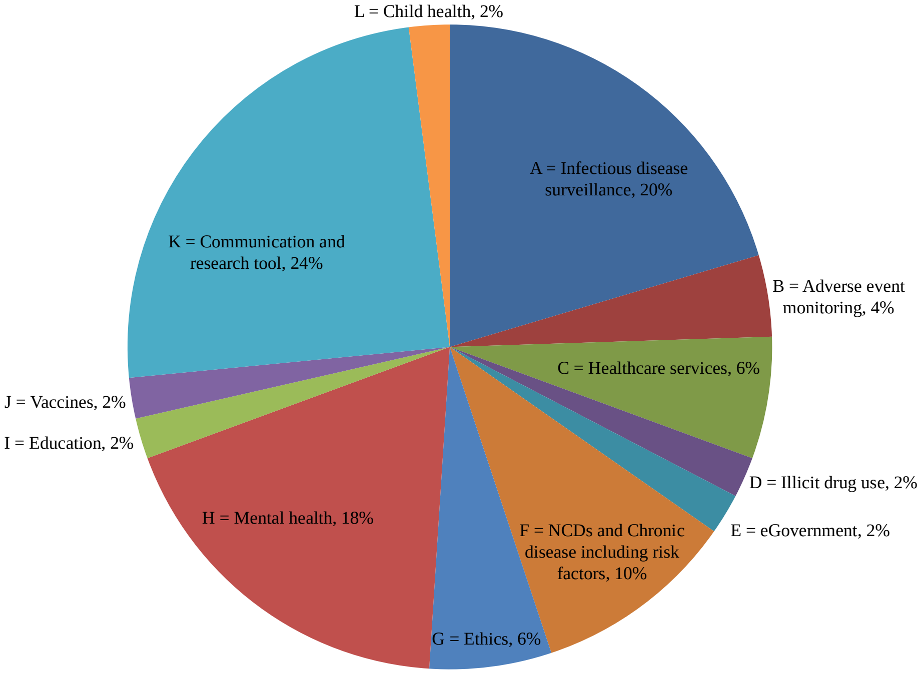
#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
		presentation in included studies; and (4) explore theoretical frameworks underpinning self-presentation in included studies.		• Significant associations were also yielded for all of the big five personality variables and narcissism. The clearest trends—based on the number of times significant associations were yielded across included studies—were as follows: (1) inauthentic self-presentation was consistently associated with low self-esteem and elevated levels of social anxiety; (2) inauthentic self-presentation was consistently more likely to occur in people high in neuroticism and narcissism; and (3) authentic/positive self-presentation was consistently associated with increased levels of self-esteem and perceived social support.			
47	Whitaker, C., Stevelink, S., & Fear, N. (2017). The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. Journal Of Medical Internet Research, 19(8), E290. [84]	To systematically review the literature regarding the current use and success of Facebook to recruit participants for health research purposes.	6 (35)	• There is growing evidence to suggest that Facebook is a useful recruitment tool and its use, therefore, should be considered when implementing future health research. • When compared with traditional recruitment methods (print,	• One social media platform included. • No protocol available. • No assessment of bias conducted.		K

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
				radio, television, and email), benefits include reduced costs, shorter recruitment periods, better representation, and improved participant selection in young and hard-to-reach demographics.			
48	Wongkoblap, A., Vadillo, M., & Curcin, V. (2017). Researching Mental Health Disorders in the Era of Social Media: Systematic Review. Journal Of Medical Internet Research, 19(6), E228. [61]	To explore the scope and limits of cutting-edge techniques that researchers are using for predictive analytics in mental health and to review associated issues, such as ethical concerns, in this area of research.	4 (48)	• Despite an increasing number of studies investigating mental health issues using social network data, some common problems persist. • Assembling large, high-quality datasets of social media users with mental disorder is problematic, not only due to biases associated with the collection methods, but also with regard to managing consent and selecting appropriate analytics techniques.	• No protocol available. • No assessment of bias.	Y	H
49	Yan, S.J., Chughtai, A.A., & Macintyre, C.R., (2017). Utility and potential of rapid epidemic intelligence from internet-based sources. International Journal of Infectious Diseases, 63, 77-87. [48]	To identify and summarise the types of Internet-based surveillance methods studied in recent literature and to identify and summarise the timeliness and accuracy outcomes of Internet-based	5 (84)	• Studies used search queries, social media posts and approaches derived from existing Internet-based systems for early epidemic alerts and real-time monitoring.	• Did not investigate variations between different health topics or conditions. • No protocol available.	Y	A

#	Systematic or Quasi-systematic Scoping Review	Review Aim/Objective/Main Focus	Number of Databases searched (Number of studies identified)	Key findings	Limitations of review	Secondary use of social media data	Topic
		methods described in the literature.		• Most studies noted improved timeliness compared to official reporting. • Internet-based methods showed variable correlation strength with official datasets, with some methods showing reasonable accuracy.	• No assessment of bias.		

The percentages of systematic, quasi-systematic and scoping reviews grouped by topic are represented in Figure 3. The majority of the 49 reviews identified focused on social media use as a communication and research tool for public health research (24%, n=12), its use for infectious disease surveillance (20%, n=10) and research into mental health topics (18%, n=9). The number of databases and search engines included in these reviews range from one to 16, whilst the number of studies identified range from two to 348, with only eight of the 49 reviews referencing a pre-defined research protocol.

Figure 3. Percentage of systematic, quasi-systematic and scoping reviews grouped by topic



Of the 24 reviews that investigated the secondary use of social media for health research, 38% (n=9) were for infectious disease surveillance, 13% (n=3) for healthcare services, 13% (n=3) as a communication and research tool and 13% (n=3) for mental health. This, therefore, suggests a gap in the existing literature associated with the secondary use of social media data to investigate topics such as NCDs and chronic disease (8%, n=2), illicit drug use (4%, n=1), adverse event monitoring (4%, n=1), eGovernment (4%, n=1), vaccines (0%) and child health (0%), as well as the ethics of collecting and analyzing this data (4%, n=1).

The use of social media as a communication and research tool

The systematic, quasi-systematic and scoping reviews that focused on social media use as a communication and research tool for health research included Capurro et al.'s review of the use of social networking sites for public health practice and research [76]. The search strategy for this review did not include discussion forums, blogs or health specific social media platforms and social networking and social media were considered synonymous. Despite this, the findings highlight that most studies identified were observational, targeted hard-to-reach populations and addressed taboo public health topics such as sexual health and mental illness.

A subsequent review, conducted by Sinnenberg et al. into the use of Twitter as a tool for health research, identified 137 eligible studies [83]. Findings highlighted that Twitter is primarily used for content analysis (56%), surveillance (26%), engagement (14%), recruitment (7%), intervention (7%) and network analysis (4%). Recruitment of study participants is further investigated in Whitaker et al.'s systematic review of the use of Facebook in recruiting participants for health research purposes, where benefits identified include reduced costs, shorter recruitment periods, better representation and improved participant selection [84].

In Moorhead et al.'s [75] systematic review of the uses, benefits and limitations of social media for health communication, 98 research articles were extracted from 10 electronic databases and four main themes (i.e. Profile, Uses, Benefits and Limitations) were then used to synthesise the results of the studies. Findings indicate that social media is used in health communications: to provide health information on a range of conditions; to provide answers to medical questions; to present information in modes other than text; to facilitate dialogue; to provide access to tailored resources; to share and collect data on experiences; as a health intervention; to deliver promotional and educational information; and to reduce the stigma associated with certain health topics. The benefits of using social media for health communications are identified as enabling an increased number of interactions with patients, generating more available health information, enabling access to tailored resources, widening the access to information and providing peer, social and emotional support and the potential for public health surveillance. A number of limitations were, however, identified, including concerns about the quality of information available, a lack of reliable health information and confidentiality and privacy issues, as well as an information overload created by the volume of data available. Moorhead et al. also identified that certain social media seem to be more effective at encouraging health behavioural changes than others. The relative immaturity of this research field is highlighted by the results of the risk of bias assessment, which infer that the majority of studies are low in quality due to their limited methodologies and exploratory nature. This is further supported by Hu et al.'s review of health communication in the digital age, where only two of the 384 studies identified investigated social media [78].

The systematic, quasi-systematic and scoping reviews discussed focus on the use of social media as a communication and research tool; however, they do not provide insight into how members of the healthcare ecosystem use these platforms. In Hamm et al's comprehensive scoping review into uses of social media by patients and caregivers, 11 databases were searched [69]. Findings highlight that discussion forums and online support groups dominate the literature, followed by social networking sites and blogs or micro-blogs. The most commonly intended use of social media is for self-care, whilst the largest proportion of conditions fall under lifestyle and weight loss and cancer with Facebook, YouTube, Twitter, MySpace, PatientsLikeMe and SecondLife being the most commonly studied social media.

Similarly, findings from Hamm et al.'s patients and HCP review [89] conclude that discussion forums are the most commonly studied social media and that the most common platforms referred to in the literature are Facebook, YouTube, Twitter, Wikipedia and SecondLife. Education is the most common user setting, with the most common objective being communication, followed by knowledge and skills.

The use of social media for public health surveillance

As introduced in Section 1.3, the terms infodemiology and infoveillance describe the automated and continuous analysis of unstructured, free text information available on the Internet, for the primary purpose of public health surveillance. The first identified systematic, quasi-systematic or scoping review to refer to infodemiology using data from social media was published in 2012 and lacked the design rigour of a full-scale systematic review. It focused on influenza-like-illness monitoring using data from Twitter [41] and was followed, in 2013, by Bernardo et al.'s scoping review into the use of internet search queries and social media for disease surveillance, which identified the first empirical study as being published in 2006 and describes how techniques for exploiting this data are evolving to increase the accuracy of signal detection [42]. Although they are insightful in terms of understanding the chronologies of this form of innovation, the findings are limited by the search of only one database and the focus on the surveillance of foodborne illness. A 2014 systematic review by Velasco et al. went further, concluding that existing surveillance systems are limited and there is a need for automated technologies to monitor health-related information on the Internet [43]. The focus on influenza-like and foodborne illnesses was later expanded to include dengue, HIV/AIDS, malaria, rabies and tuberculosis as part of Zhou et al.'s 2016 review [46].

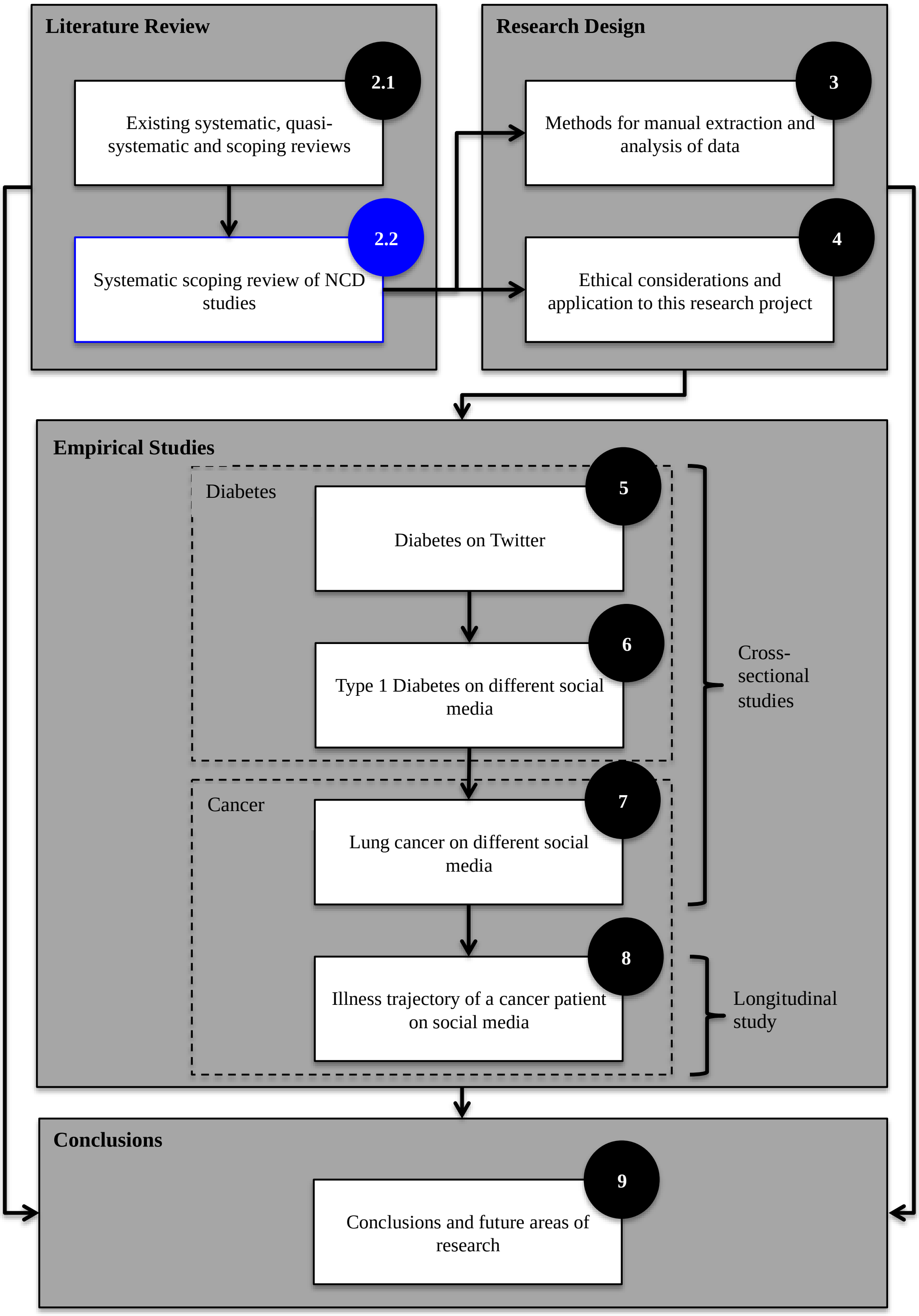
O'Shea et al.'s systematic review describes a wide variety of technologies and data sources for gathering, processing and disseminating data for event-based internet biosurveillance systems [47], whilst Yan, Chughtai and Macintyre's review investigated the timeliness and accuracy of the methods used [48]. Charles-Smith et al.'s review goes further by highlighting the challenges of translating retrospective research data into predictive practice for actionable disease surveillance and outbreak management [44]. These reviews, alongside the others listed, focus predominantly on the secondary use of social media data for the surveillance of infectious disease.

Five reviews were identified that investigate NCDs and chronic disease. The first of these was published in 2013 and focused on the risks associated with alcohol consumption for people with Type 1 diabetes [49]. Although limited in terms of sample size, the six studies identified all sourced data from social media. Also limited in terms of the number of databases searched (1) and the studies identified (9) was Cotter et al.'s systematic review into the effectiveness of internet-based interventions on the management of Type 2 diabetes [50]. A systematic review on social media for chronic disease was later published in 2015 [86] and focused on understanding the clinical outcomes associated with using such technologies for patient support, education and disease management across different conditions. This review, however, was also limited by its search of only one database.

Unhealthy diets are a known risk factor linked to NCDs. A search of 10 databases by Nour et al. identified 11 studies and demonstrated the efficacy of social media-based behavioural interventions on nutritional knowledge and attitudes [51]. This is supported by evidence from Rose et al.'s review of eight databases and 26 studies, which also included exercise and physical activity [52]. Despite this, none of the NCD and chronic disease reviews identified went so far as to examine the type of methods or tools used to extract and analyse data from social media, the academic discipline, the setting of the research or the ethical issues considered.

With respect to the latter, a quasi-systematic review, published by Conway in 2014, produced a taxonomy of ethical concepts that were identified from studies using Twitter for public health surveillance and research [73]. The systematic review by Golder et al. went further and analysed studies reporting people's attitudes to research sourcing data from social media. This revealed a wide variation in attitudes, from the very positive to the very concerned, depending on the purpose and quality of the research, researcher affiliation, the potential for harm and the methods used [72]. Although it used an impressive 16 databases, this review did not examine regional, disciplinary, condition or health-topic-specific variations and the authors note that the demographic characteristics of respondents were unclear in most studies.

The rising popularity of forms of social media, since their inception around 20 years ago, has been echoed in the growth of health-related research using data extracted and analysed from them. This has created a demand for further literature reviews, to synthesise and understand the emerging evidence-base and inform future activities. Existing reviews, such as those described earlier in this chapter, tend to be narrow in scope, with limited consideration of the different types of data, analytical methods and ethical issues involved. There has also been a tendency for research to be siloed within different academic communities (e.g. computer science, public health), hindering knowledge translation.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

2.2 Status of existing multi-disciplinary research literature that analyses data posted on social media for health research

To address the limitations of existing reviews discussed in Section 2.1, a comprehensive, systematic, scoping review was undertaken to systematically capture the broad corpus of published, health-related research sourcing data from social media. This chapter describes the methods, results, discussion and implications of this comprehensive, systematic, scoping review of NCD-related research using social media data. The protocol was published in the *BMJ Open* journal (Appendix A.2.3).

2.2.1 Methods

A scoping review is a type of quasi-systematic review that is increasingly used to understand research on emerging innovations, which may be poorly indexed, distributed across published and grey literature or located in different academic disciplines [90-92]. Scoping reviews typically follow Arksey and O'Malley's five key stages: (1) identifying the research question; (2) identifying the relevant studies; (3) selecting the studies; (4) charting the data; and (5) collating, summarising and reporting the results. Developing *a priori* review protocol is useful for managing this complexity, while formative research can aid the design of such protocols by identifying relevant terminologies, topics and evidence sources. Scoping reviews are mainly aimed at mapping the evidence landscape rather than establishing the effectiveness of particular interventions and, typically, do not involve critical appraisal of study methodology or detailed extraction of outcomes data [93].

Stage 1: identifying the research question

In addition to the overarching review objective to capture and profile a broad corpus of published multi-disciplinary research in which data sourced from social media have been used to monitor, understand or evaluate aspects of health and disease, several specific questions were used to guide the analysis of existing research evidence, as listed in Table 3.

Table 3. Research questions that guide the systematic scoping review

Aspect	List of questions
General	What is the total number of studies published by year?
	What terms are being used to describe the nature of this research?
	Which academic communities are most active in health research using data from social media?
	Where are study authors located, according to their affiliation?
	What is the geographical scope of the social media data analysed in these studies?
	For what purposes are social media data being used in this research?
Topic	Which health topics and NCD are being studied?
Social Media Type	Which social media platforms or sites are being used as sources of data?
Extract and analysis	How large is the sample size and what units of analysis are being applied?
	How are data from social media being extracted and analysed and which proprietary tools are being used?
Ethics	How are ethical considerations applied in the published research?

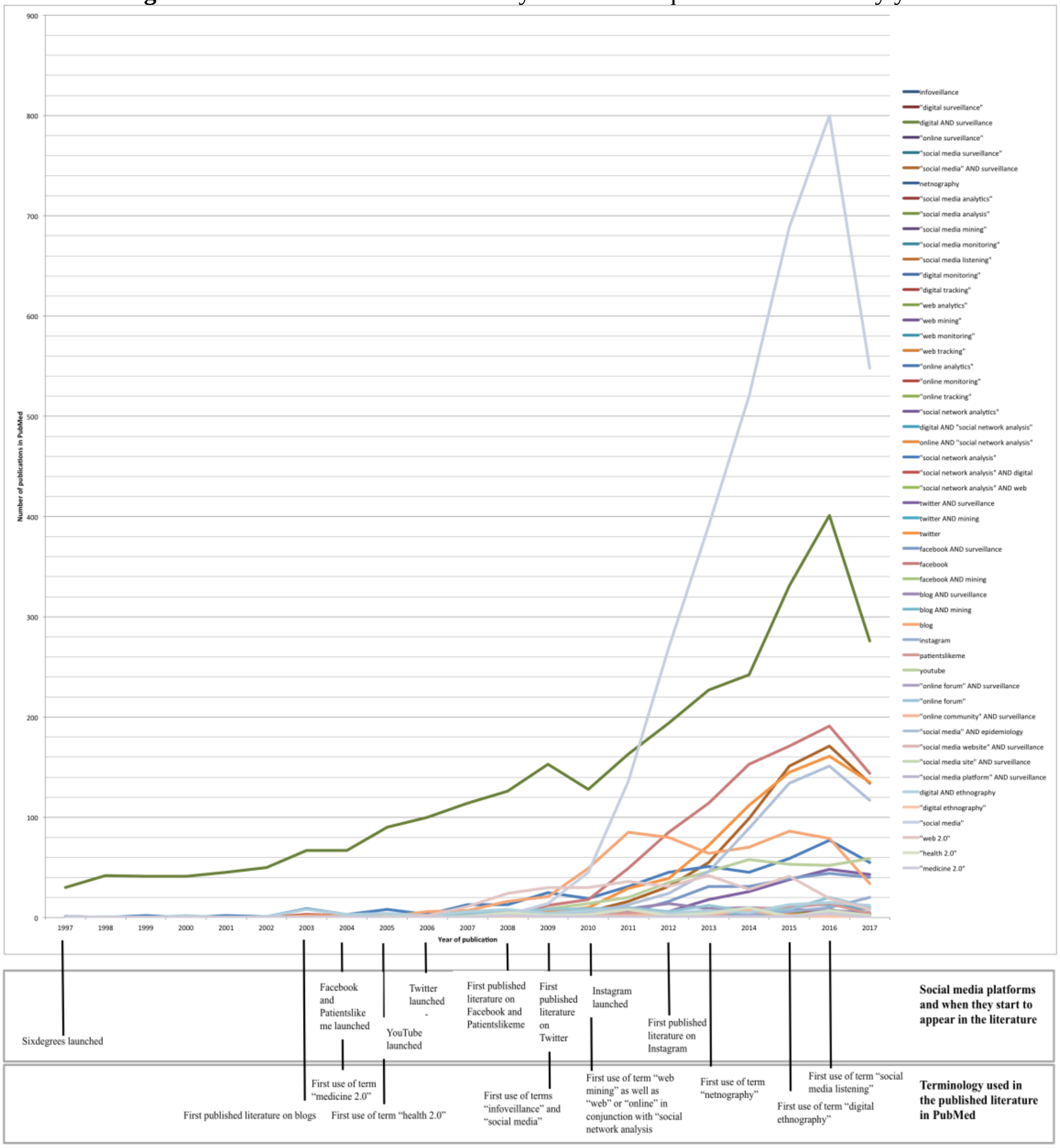
Stage 2: identifying relevant literature

Comprehensive scoping reviews aim to capture literature from a range of electronic databases. As such, the approach for this review included a systematic search of peer-reviewed studies using five health, informatics, business and cross-disciplinary electronic databases: Association of Computer Machinery (ACM), Applied Social Sciences Index and Abstracts (ASSIA), Institute of Electrical and Electronics Engineers (IEEE), PubMed and Scopus. The search strategy for each of the databases was defined in consultation with a senior librarian. The five electronic databases took into account time and funding constraints, although the sources targeted are likely to capture most of the relevant literature [94].

Given the breadth and changing popularity of terms related to social media and social media mining, a formative analysis was undertaken to understand those most likely to yield articles relevant to the review objectives. An extensive list of terms was created, based on the search strings specified in a recent systematic review of social media in the context of e-government in public health [70], supplemented with terms captured from iterative online searches. The list included generic terms related to social media (e.g. digital, blog, social network, Web 2.0), named social media (e.g. Twitter, Facebook), types of social media mining (e.g. surveillance, scraping, listening, infoveillance) and analytics (e.g. netnography, social network analysis) [see Appendix A.3 for the full list]. Separate search queries for each term were run in PubMed, constructed as [term] AND (health OR illness OR disease) AND human. Searches yielding one or more hits were charted over time, to examine their changing frequency in the research literature. Out of the 72 terms tested in PubMed, 52 produced hits. The timeline was cross-referenced with key historical events, such as the introduction of new platforms or new methodologies. The results are shown in Figure 4.

As can be seen from Figure 4, 91% of relevant studies were published in the past nine years (11,154 results between 2009-2017, vs. 1,048 between 1997-2008). Not surprisingly, the search term ‘social media’ produced the highest number of results overall compared to other search queries. Regarding specific social media, searches with sixdegrees and Linkedin produced no results, compared with other social media such as Facebook (1,152 hits, between 2008-2017), Twitter (939 hits, between 2009-2017), YouTube (351 hits, between 2008-2017), PatientsLikeMe (60 hits, between 2008-2017) and Instagram (40 hits, between 2012-2017). A time lag of four years was observed between when Facebook, PatientsLikeMe and Twitter were launched in 2004, 2004 and 2006 respectively and the first research article was published, while the equivalent latency for YouTube and Instagram was two years. The individual search terms ‘surveillance’ (665), ‘epidemiology’ (581) and ‘ethnography’ (110) produced the highest yields, compared with terms specific to digital research methods. With respect to the latter, temporal changes in the use of terms such as ‘infoveillance’, ‘netnography’, ‘social media listening’, ‘social media analytics’ and ‘social media mining’ indicate the evolution of innovations and research perspectives; however, including such specialist terms was not critical to identifying relevant studies.

Figure 4. Results of the formative analysis of search queries in PubMed by year



As previously noted, a systematic search of five health, informatics, business and cross-disciplinary electronic databases – ACM, ASSIA, IEEE, PubMed and Scopus – was undertaken to identify relevant peer-reviewed studies. Table 4 describes the search queries used to interrogate these sources. These take account of insights gathered during the formative analysis, as well as the search functionality of each of the databases.

Table 4. Search query applied to each electronic database

Name of electronic database	**Search query**
ACM	("social media" OR "blog" OR "microblog" OR "wiki" OR "virtual world" OR "discussion forum" OR "online community") AND ("surveillance" OR "epidemiology" OR "infoveillance" OR "ethnography" OR "netnography" OR "mining") AND ("health" OR "disease" OR "illness") ACM Full Text Collection
ASSIA	("social media" OR "blog" OR "microblog" OR "wiki" OR "virtual world" OR "discussion forum" OR "online community") AND ("surveillance" OR "epidemiology" OR "infoveillance" OR "ethnography" OR "netnography" OR "mining") AND ("health" OR "disease" OR "illness") Language: English Publication date: 1997-2017 Document type: Article
IEEE	("social media" OR "blog" OR "microblog" OR "wiki" OR "virtual world" OR "discussion forum" OR "online community") AND ("surveillance" OR "epidemiology" OR "infoveillance" OR "ethnography" OR "netnography" OR "mining") AND ("health" OR "disease" OR "illness") Included: Conference, Journals Excluded: Courses Publication date: 1997-2017 Full text and metadata search
PubMed	((((((facebook OR twitter OR youtube OR Instagram OR Pinterest OR Linkedin OR six degrees OR patientslikeme OR Mumsnet OR "social media" OR "blog" OR "microblog" OR "wiki" OR "virtual world" OR "discussion forum" OR "online community") AND ("surveillance" OR "epidemiology" OR "infoveillance" OR "ethnography" OR "netnography" OR "mining") AND ("health" OR "disease" OR "illness"))) AND ("1997/01/01"[PDat] : "2017/12/31"[PDat]) AND Humans[Mesh])) Article type: Journal Article Language: English
Scopus	("social media" OR "blog" OR "microblog" OR "wiki" OR "virtual world" OR "discussion forum" OR "online community") AND ("health" OR "disease" OR "illness") Article type: Article, Conference Paper Publication date: 1997-2017

Stage 3: study selection

A systematic search of the electronic databases was conducted and the titles and abstracts screened based on the pre-defined inclusion and exclusion criteria. Peer-reviewed journal articles and conference papers that were included describe empirical research using data from social media such as Twitter or Facebook, whether extracted or studied *in situ*, using either manual or automated methods. Studies not in English, dissertations/theses, reports or abstracts, letters to the editor and feature articles and articles intended as marketing or advertising material were excluded. A publication timeframe of 1997-2017 was applied. See Box 1 for the inclusion and exclusion criteria for the scoping review.

Box 1. Article inclusion and exclusion criteria for the systematic scoping review

Inclusion

Types of publication: Peer-reviewed research articles. Full conference papers.
Language: English
Publication timeframe: 1997-2017
Types of research: Empirical studies using health-related text data from social media, extracted or studied *in situ*, using both manual and automated methods.

Exclusion

Types of publication: Dissertations/theses; Reports or abstracts only; Letters to the editor; Marketing or advertising material; Reviews or editorials
Language: Not English
Types of research: Studies based on text and non-text data (e.g. GPS, rating, videos, images) from online sources other than social media (e.g. internet search histories, online news reports).
Commercial research aimed at obtaining market intelligence or informing product promotion.
Studies examining social media platforms, rather than using them as a source of data. Studies describing social media as a communication or broadcasting channel (e.g. for public health promotion).

The four-phase PRISMA statement for identification, screening, eligibility and inclusion of the literature was applied [95]. The PRISMA flowchart (Figure 6) reflects each phase of the process.

EndNote reference management software was used to manage the records retrieved from searches. The generated citations were screened with the help of EPPI-Reviewer 4 systematic review software and the classification was conducted in Microsoft Excel.

Stage 4: charting the data

The purpose of charting data in scoping reviews is to produce a descriptive summary of the results. For this stage, existing classification frameworks that correspond with the research questions listed previously were identified. These formed the basis of the data charting form. The classification frameworks consider the purpose for which social media data are used in the research, the method of data extraction and analysis (including any automated data mining tools used), the analytical-interpretive approach used (including stated theoretical perspectives), the locus of the research by academic institution and geographical scope of the data, the academic discipline associated with the research as well as whether and how ethical issues or guidelines are considered. Each of these existing classification frameworks and taxonomies are described in Table 5 and further detailed later in this chapter, in relation to the specific research questions that guide the systematic scoping review. These classification frameworks were not intended to be totally prescriptive and additional emerging themes were captured throughout the forthcoming analysis. These themes were later used to identify gaps and inconsistencies represented in the existing frameworks, for future consideration and refinement.

Table 5. Existing classification frameworks that correspond with the research questions for the systematic scoping review

Research question	**Existing classification framework to be applied**	**How the classification framework will be applied**
What is the total number of studies published by year?	Not specified.	The year that the eligible study was published.
What terms are being used to describe the nature of the research?	Not specified.	The terminology used by study authors to describe the aims or methodologies used in their study will be captured e.g. infoveillance, netnography, social listening.
Which academic communities are most active in conducting health research studies that source data from social media?	The 27 Scimago Journal Subject Areas [96] including medicine, computer science, health professions, business management and accounting, as well as social sciences.	The journal in which the eligible study is published, along with the affiliation of all authors as well as any sources of funding for the study (if shown) will be classified according to the disciplinary taxonomy used by the Scimago Journal ranking portal. These categories are not mutually exclusive.
Where is the affiliation of the author located?	World Bank Regional and Lending Groups for Countries, which includes 195 countries and four lending groups [97].	The geographical location of the author conducting the eligible study will be classified according to the country and lending groups

Research question	Existing classification framework to be applied	How the classification framework will be applied
		provided by the World Bank for 2017. These categories are not considered to be mutually exclusive.
What is the geographical scope of the sample of social media data analysed in the study?	World Bank Regional and Lending Groups for Countries, which includes four lending groups [97].	The geographical location of the population sample analysed within the study will be classified according to the regional and lending groups provided by the World Bank for 2017. These categories are not mutually exclusive.
What is the purpose for which social media data are being used in the research?	Not specified.	The purpose of the eligible study will be captured.
Which NCD is being investigated?	NCDs as defined by the World Health Organization [34] and excluding mental health conditions.	The type of condition being studied in the eligible study is classified according to the list of NCDs. These categories are not considered to be mutually exclusive.
Which social media are used as a source of data?	Kaplan and Haenlein's eight types of social media [5].	The type of social media from which the data for the eligible study was sourced will be classified according to Kaplan and Haenlein's eight types of social media. The name of the social media will also be captured.
How large is the sample size and what units of analysis are being applied?	Not specified.	The sample size and unit of analysis of the eligible study will be captured.
How is the data from social media extracted and analysed and which proprietary tools are used?	If the analysis is conducted using 'automated' techniques, these will be classified according to the data mining techniques surveyed by Injadat et al.[98]	Whether the eligible studies applied 'automated' or 'manual' data extraction methods, the named type of analysis performed in the eligible study will be captured as well as any reference to proprietary tools used.
How are ethical considerations applied in the published research?	Conway's taxonomy of ethics concepts for the use of Twitter in public health surveillance and research [73], which can be applied across all manner of social media.	The application of ethical concepts in the eligible studies will be classified according to the ten high-level categories identified by Conway. These categories are not considered to be mutually exclusive.

Each of these classification frameworks are described in the subsequent sections.

Type of social media from which data is sourced

The term ‘social media’ itself is often used synonymously with ‘social networking’; however, for the purpose of this systematic scoping review, social networking will be considered to be one of the eight types of social media described in Table 1 [5, 15]. Only studies that source text data from social media were included in the systematic scoping review. Social bookmarking, content communities and virtual worlds were excluded unless text content was analysed in the studies.

Research discipline contributing to the study

Given the highly multi-disciplinary nature of health research, the author affiliation, the journal or conference proceedings in which the study is published and the source of funding for the study was classified into subject areas according to the taxonomy provided by the Scimago Journal ranking portal Box 2.

Box 2. Scimago Journal Subject Areas

1. Agriculture and biological sciences
2. Arts and humanities
3. Biochemistry, genetics and molecular biology
4. Business, management and accounting
5. Chemical engineering
6. Chemistry
7. Computer science
8. Decision sciences
9. Dentistry
10. Earth and planetary sciences
11. Economics, econometrics and finance
12. Energy
13. Engineering
14. Environmental science
15. Health professions
16. Immunology and microbiology
17. Materials science
18. Mathematics
19. Medicine
20. Multi-disciplinary
21. Neuroscience
22. Nursing
23. Pharmacology, toxicology and pharmaceutics
24. Physics and astronomy
25. Psychology
26. Social sciences
27. Veterinary

Location of the study and population sample

The location of the study and the population sample were classified according to the country and lending groups provided by the World Bank in 2017 [97] (Box 3).

Box 3. World Bank Country and Lending Groups

Low-income economies	Lower-middle-income economies
Afghanistan	Angola
Benin	Bangladesh
Burkina Faso	Bhutan
Burundi	Bolivia
Central African Republic	Cabo Verde
Chad	Cambodia
Congo, Dem. Rep.	Cameroon
Eritrea	Comoros

Ethiopia	Congo, Rep.
Gambia, The	Côte d'Ivoire
Guinea	Djibouti
Guinea-Bissau	Egypt, Arab Rep.
Haiti	El Salvador
Korea, Dem. People's Rep.	Eswatini
Liberia	Ghana
Madagascar	Honduras
Malawi	India
Mali	Indonesia
Mozambique	Kenya
Nepal	Kiribati
Niger	Kyrgyz Republic
Rwanda	Lao PDR
Sierra Leone	Lesotho
Somalia	Mauritania
South Sudan	Micronesia, Fed. Sts.
Syrian Arab Republic	Moldova
Tajikistan	Mongolia
Tanzania	Morocco
Togo	Myanmar
Uganda	Nicaragua
Yemen, Rep.	Nigeria
	Pakistan
	Papua New Guinea
	Philippines
	São Tomé and Principe
	Senegal
	Solomon Islands
	Sudan
	Timor-Leste
	Tunisia
	Ukraine
	Uzbekistan
	Vanuatu
	Vietnam
	West Bank and Gaza
	Zambia
	Zimbabwe
Upper-middle-income economies	**High-income economies**
Albania	Andorra
Algeria	Antigua and Barbuda
American Samoa	Aruba
Argentina	Australia
Armenia	Austria
Azerbaijan	Bahamas, The
Belarus	Bahrain
Belize	Barbados
Bosnia and Herzegovina	Belgium
Botswana	Bermuda
Brazil	British Virgin Islands
Bulgaria	Brunei Darussalam
China	Canada
Colombia	Cayman Islands
Costa Rica	Channel Islands
Cuba	Chile
Dominica	Croatia
Dominican Republic	Curaçao
Ecuador	Cyprus
Equatorial Guinea	Czech Republic
Fiji	Denmark

Gabon Georgia Grenada Guatemala Guyana Iran, Islamic Rep. Iraq Jamaica Jordan Kazakhstan Kosovo Lebanon Libya Malaysia Maldives Marshall Islands Mauritius Mexico Montenegro Namibia Nauru North Macedonia Paraguay Peru Romania Russian Federation Samoa Serbia South Africa Sri Lanka St. Lucia St. Vincent and the Grenadines Suriname Thailand Tonga Turkey Turkmenistan Tuvalu Venezuela, RB	Estonia Faroe Islands Finland France French Polynesia Germany Gibraltar Greece Greenland Guam Hong Kong SAR, China Hungary Iceland Ireland Isle of Man Israel Italy Japan Korea, Rep. Kuwait Latvia Liechtenstein Lithuania Luxembourg Macao SAR, China Malta Monaco Netherlands New Caledonia New Zealand Northern Mariana Islands Norway Oman Palau Panama Poland Portugal Puerto Rico Qatar San Marino Saudi Arabia Seychelles Singapore Sint Maarten (Dutch part) Slovak Republic Slovenia Spain St. Kitts and Nevis St. Martin (French part) Sweden Switzerland Taiwan, China Trinidad and Tobago Turks and Caicos Islands United Arab Emirates United Kingdom United States Uruguay Virgin Islands (U.S.)

Ethical considerations applied

Conway [99] suggests a taxonomy of ethical concepts relevant to the use of Twitter for public health surveillance and research. This taxonomy consists of ten high-level categories:

1. Privacy
2. Informed consent
3. Ethical theory
4. Institutional review board (IRB)/regulation
5. Traditional research versus social media (e.g. Twitter) research
6. Geographical information
7. Researcher lurking
8. Economic value of personal information
9. Medical exceptionalism
10. Benefit of identifying socially harmful medical conditions.

Although originally intended for Twitter-based public health research, these ethical concepts can be applied across all manner of social media.

Automated data extraction technique applied

Injadat et al. [98] conducted a survey of data mining techniques applied by researchers in the area of social media and identified 19 different techniques:

1. AdaBoost
2. Artificial Neural Network
3. Apriori
4. Bayesian Networks
5. Decision Trees
6. Density Based Algorithm
7. Fuzzy
8. Genetic Algorithm
9. Hierarchical Clustering
10. K-Means
11. k-nearest Neighbors
12. Linear Discriminant Analysis

13. Linear-Regression
14. Logistic Regression
15. Markov
16. Maximum Entropy
17. Novel
18. Support Vector Machine
19. Wrapper.

Classification of NCDs

NCDs are defined by the WHO as those which do not pass from person to person, but are commonly linked to behavioural or physiological risk factors such as tobacco use, physical inactivity, alcohol consumption and unhealthy diet [34]. NCDs are often referred to as chronic diseases and include cardiovascular diseases, cancers, chronic respiratory diseases and diabetes.

For the purposes of this thesis and due to the significant volume of literature identified during the screening phase, the decision was made to diverge from the original published protocol and focus on studies related to NCDs during the inclusion phase. The full corpus of literature will be included in future analysis and submitted for consideration for publication in a journal.

Stage 5: collating, summarising and reporting the results

The extracted data was tabulated, with rows relating to articles, columns to classification variables and cells containing the relevant information. Both frequency analysis and trend analysis were used to chart the classified results.

Using frequency analysis, the counts and percentages of eligible studies were calculated. Studies were grouped based on the classification frameworks and taxonomies applied. The result of this analysis is a map of eligible studies represented in bubble plots, graph or tabular form.

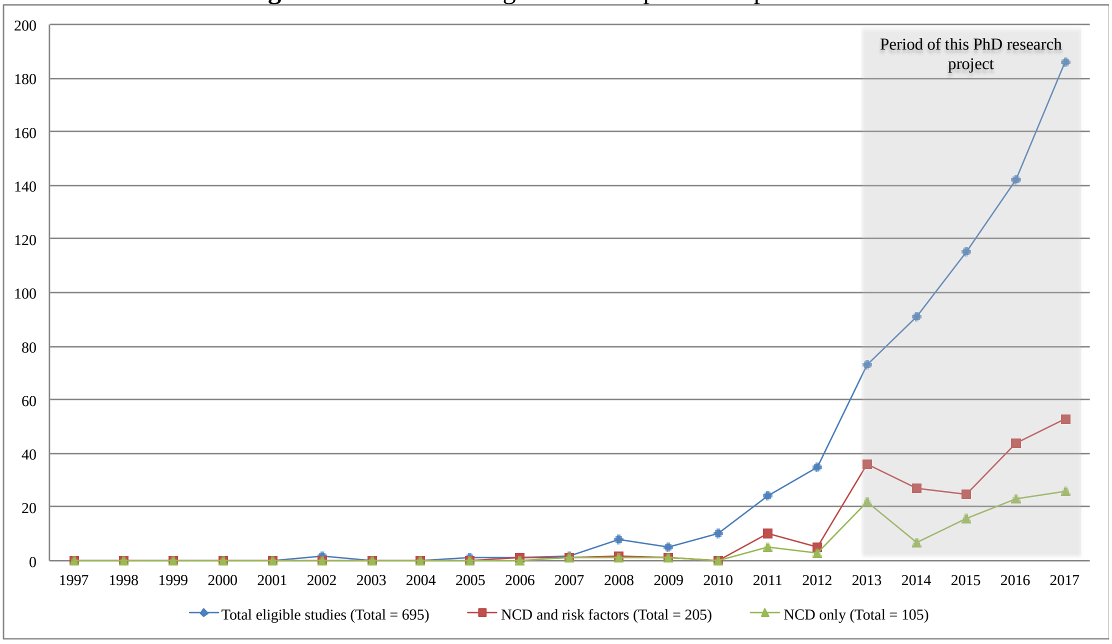
Trend analysis were used to present the changing frequency of research over time, based on the aforementioned classification criteria. A map showing the characteristics of included studies are presented in this thesis using static graphs.

Although some scoping reviews include a consultation phase, this one focused on published research and online sources already in the public domain. No ethics approval was therefore required for the conduct of this systematic scoping review. The completed scoping review will be submitted for publication to a peer-reviewed, inter-disciplinary Open Access journal, in addition to conferences on public health and digital research. Findings will be presented using both static and multimedia visualisation tools.

2.2.2 Results

The search of the five databases returned 13,008 articles. After screening of the full-text this yielded 695 articles that met the inclusion criteria, 205 of these were relevant to NCDs and risk factors, whilst 105 were associated with NCDs only. In Figure 5, the number of studies published per year from the corpus of 695 eligible studies is compared to those associated with NCDs. Figure 6 documents the review process using a PRISMA flow chart.

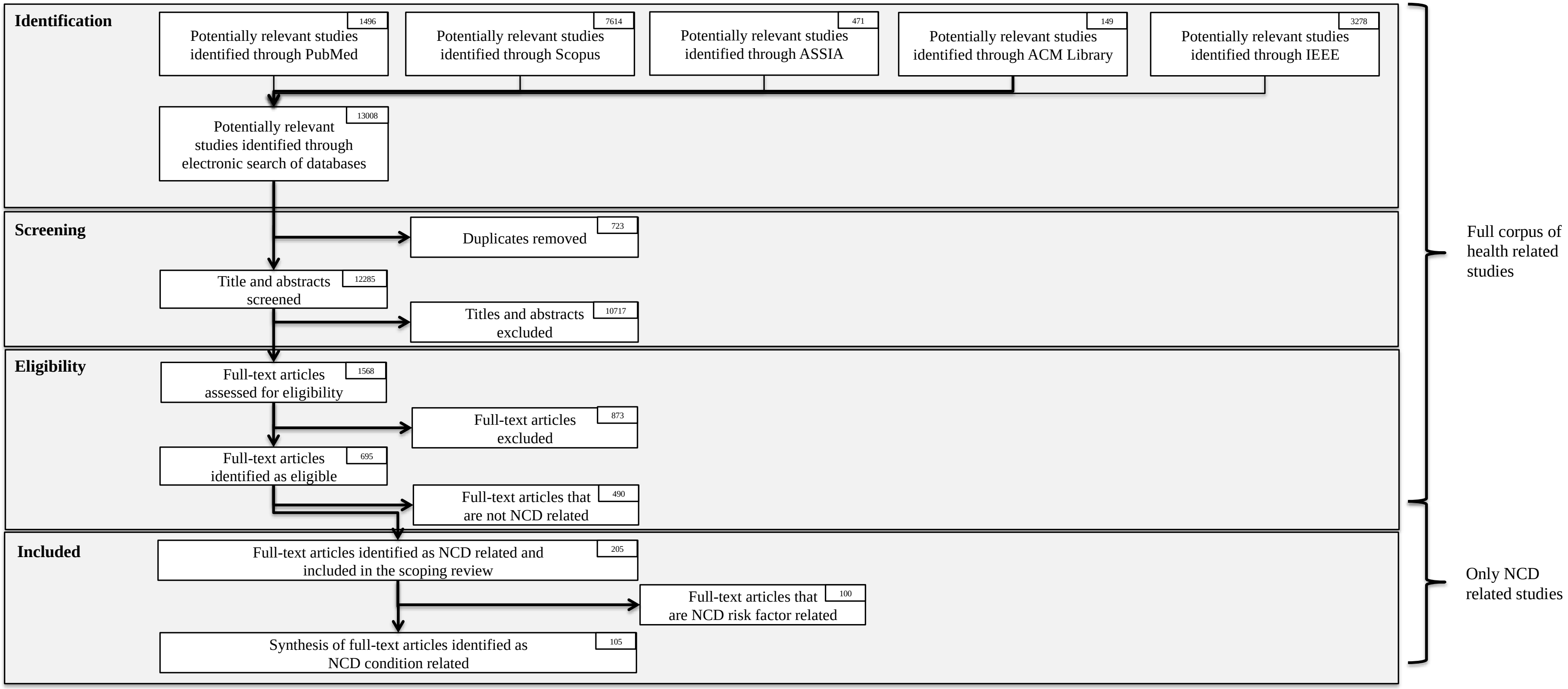
Figure 5. Number of eligible studies published per annum



The results of this systematic scoping review highlight a rapid increase in the number of studies published in the last five years, with 87% of total empirical studies published between 2013 and 2017. The number of NCD-related studies has also increased over this period, albeit at a slower rate, a finding that highlights the continued lack of empirical studies sourcing data from social media to investigate NCDs. The increasing number of empirical studies in recent years reiterates the importance of conducting and repeating structurally, comparable reviews at regular time intervals in order to draw conclusions about the evolution of the field.

Contrary to more traditional, sequential approaches to PhD research projects where a review of the literature is conducted at the start of the PhD programme, initial, semi-structured, informal systematic reviews were conducted in 2013 and 2015 as part of this research project. These were used as a means of informing the direction for the empirical studies described in later chapters. The structured, systematic scoping review described in this chapter was conducted in the latter part of the 6-year research period, in order to reflect the status of existing multi-disciplinary research literature in 2018.

Figure 6. PRISMA flowchart for systematic scoping review of the literature



Studies that were excluded during the screening and eligibility phases (Table 6) were excluded primarily due to the research design and subject area of the study, if the type of publication was not a journal article or conference proceeding, if the language in which the study manuscript was written was not English, if the date of publication was outside the 2007-2017 time-frame or if no full-text article was available.

Table 6. Reason for study exclusion during screen and eligibility phases of the systematic scoping review

Phase	Reason for exclusion	Number of studies excluded
Screening of Titles and Abstracts (Total number of studies screened = 12285)	Type of study based on research design or subject area	10028
	Type of publication	369
	Language in which the study is written	17
	Publication date outside the 2007-2017 time-frame	8
	Duplicate studies not identified or removed using EndNote and EPPI Reviewer duplication search functionality	296
Full-text articles assessed for eligibility (Total number of studies assessed = 695)	Type of study based on research design or subject area	572
	No full-text article available	300

Studies that were excluded due to research design or subject area were those that: did not relate directly to human health topics; were not empirical studies that extracted and analysed text content from social media and instead sourced data such as GPS data, number of followers, videos and images; were systematic, quasi-systematic or scoping reviews (Section 2.1).

By searching the titles and abstracts of the 695 eligible studies using EPPI-Reviewer 4, a preliminary analysis by topic (Figure 7) revealed that most of the empirical research that analyses data from social media, investigates topics outside the selected search terms i.e. ‘other’ (13.0%) followed by drug (9.6%), cancer (7.7%), mental (6.1%), flu* (5.4%) and infect* (5.1%). This is somewhat contradictory to the earlier results and therefore warrants further investigation as part of the full scoping review described in the protocol.

Figure 7. Percentage of the 695 eligible studies by topic (preliminary results)

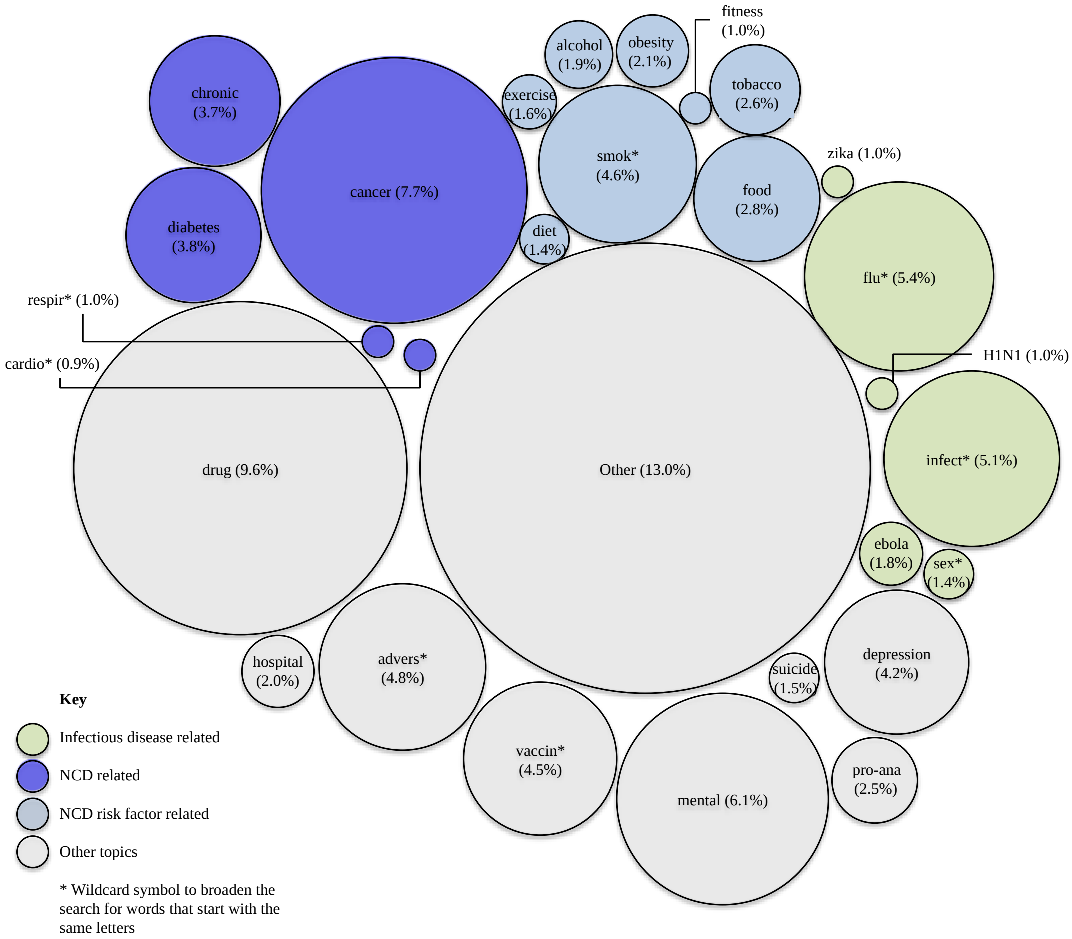


Table 7 provides a summary of the 105 eligible NCD-related studies that source data from social media. The subsequent sections describe and analyse the characteristics of these studies.

Table 7. The 105 eligible NCD-related studies that source data from social media

#	Study Title	Study Purpose	Key findings
1	Social Media as a Platform for Information About Diabetes Foot Care: A Study of Facebook Groups [100]	To provide insight into the general characteristics and usefulness of the information posted on diabetes foot care-related Facebook groups.	16 eligible diabetes foot care-related Facebook groups were identified with a total of 103 eligible posts. Of the total 103 timeline posts, 45.6% posts were categorized as useful, while the remaining posts were not useful. Top mentioned diabetes foot care practice was checking feet daily. Patient welfare organizations, doctors, nurses and podiatrists could use Facebook to provide support to educating diabetes patients and their caregivers by disseminating useful and authentic knowledge and information related to diabetes foot care.
2	Communicating about cancer through Facebook: A qualitative analysis of a breast cancer awareness page [101]	To identify the function and uses of a popular Facebook Page from a nonprofit organisation devoted to raising awareness about breast cancer.	The findings have potential implications for health promotion efforts using social media platforms. Common themes and characteristics are identified, including open mic communication, scarcity of health information, the commodification of breast cancer, unpredictable locations of conversation, and the use of gendered images and language.
3	A novel data-mining approach leveraging social media to monitor and respond to outcomes of diabetes drugs and treatment [102]	To transform forum posts into vectors to scan for patterns in the responses to gauge consumer opinion of the drugs and devices used to treat diabetes mellitus.	The result of the study was a word list compilation that correlates certain positive and negative word cluster groups with medical drugs and devices. The implication of this novel data-mining method could open new avenues of research into rapid data collection, feedback, and analysis that would enable improved outcomes and solutions for public health.
4	A novel data-mining platform leveraging social media to monitor outcomes of Januvia [103]	To gauge the experiences of the diabetes mellitus drug Januvia using a novel data-mining method.	The result of this study was a word list compilation correlating positive and negative word cluster groups and a web of influential users on Januvia. The implications could open new research avenues into rapid data collection, feedback, and analysis that would enable improved solutions for public health.

#	Study Title	Study Purpose	Key findings
5	Network-based modeling and intelligent data mining of social media for improving care [104]	To assess correlations between user posts and positive or negative opinion on the drug using self-organising maps.	Most of the user data was linked to positive sentiment on lung cancer treatment, therefore reflecting the general positive view of the users. The mining technique was able to identify information brokers which seemed to play important roles in the shaping the information content of the forum as well as identify potential side effects consis- tently discussed by groups of users. The consensus on Erlotinib depends on individual patient experience.
6	A novel data-mining approach leveraging social media to monitor consumer opinion of Sitagliptin [105]	To transform the posts of a forum dedicated to diabetes mellitus Type 2 into vectors to be able to intelligently mine consumer opinion of the drug Sitagliptin.	The results were a compilation of user's clusters and their correlated (positive or negative) opinion of the drug and can open new avenues of research into rapid data collection, feedback, and analysis that can enable improved outcomes and solutions for public health and important feedback for the manufacturer.
7	Health information sharing on Facebook: An exploratory study on diabetes mellitus [106]	To provide insight into the type of information sharing conducted on Facebook for diabetes mellitus in the Arabic-speaking world and the potential health consequences.	1551 diabete mellitus posts from 7 Arabic sites were analysed. There were 458 (30%) FB posts from Egypt with no posts from Somalia, Yemen, Comoros, and Djibouti. The majority of the posts, 863 (56%), were from females. The focus of the posts was on sharing personal experiences, raising awareness, providing spiritual support, sharing latest research, and providing education on diabetes mellitus. A large number of the posts by people in 40-60 year age group were around finding out diagnosis related information due to limited access to care in their home countries. Patients with diabetes mellitus are increasingly sharing their health information with other Facebook users.
8	Accuracy and readability of cardiovascular entries on Wikipedia: Are they reliable learning resources for medical students? [107]	To evaluate accuracy of content and readability level of English Wikipedia articles on cardiovascular diseases, using quality and readability tools.	This study was based on 47 English Wikipedia entries on cardiovascular diseases. The Wikipedia entries were not aimed at a medical audience and should not be used as a substitute to recommended medical resources. Course designers and students should be aware that Wikipedia

#	Study Title	Study Purpose	Key findings
			entries on cardiovascular diseases lack accuracy. Further improvement of the Wikipedia content of cardiovascular entries would be needed before they could be considered a supplementary resource.
9	Real-world experience with colorectal cancer chemotherapies: Patient web forum analysis [108]	To better understand patient experience with colorectal cancer chemotherapies in the real-world setting.	The threads included 1522 posts by 264 individuals. Demographics were identified for 83% of the posters. Of these, 83% were colorectal cancer patients and 17% were family members; 76% were females, and the mean patient age was 49 years. The majority had advanced cancer (44% stage IV or metastatic, 40% stage III). The most common themes were side effects (62.3% of posts), treatment response (13%), and impact on personal, social, and work lives, and emotional distress (23.9%). The posters came to the online forums to have an emotional outlet, share experience, and seek advice. The emotional impacts primarily exemplified resilience and positive coping strategies. Formal knowledge regarding the likelihood of treatment response, magnitude of benefit, or side effects was lacking, which lead to uncertainty and anxiety. However, patients expressed appreciation for the availability of treatment options and the hope they provide. Online colorectal cancer communities provide patients with convenient and valuable emotional support and disease information. Colorectal cancer and treatments may have profound impacts beyond efficacy and toxicity. Systematic information and decision tools may help to minimise uncertainties and help patients manage expectations and emotional distress.
10	Using social media data to understand the impact of promotional information on laypeople’s discussions: a case study of Lynch syndrome [109]	To understand how promotional Lynch syndrome–related health information impacts laypeople’s discussions on Twitter.	The most discussed themes in the 16,667 Lynch syndrome-related tweet were treatment (n=4080) and genetic testing (n=3073). The topic distributions in laypeople's discussions were found to be

#	Study Title	Study Purpose	Key findings
11			similar to the distributions in promotional Lynch syndrome-related information. When comparing monthly trends, laypeople's discussions had a strong correlation with promotional Lynch syndrome-related information on awareness, while there were moderate correlations on screening, genetic testing, treatment, and risk. It was discovered that the Colon Cancer Awareness Month (March 2017) and the Lynch Syndrome Awareness Day (March 22, 2017) had significant positive impacts on laypeople's discussions and their attitudes.
11	Co-training over Domain-independent and Domain-dependent features for sentiment analysis of an online cancer support community [110]	To analyse the sentiment of user messages on an online cancer support community.	Co-training is an effective way to combine two information with respect to sentiment classification performance. It was found that the last sentences of the posts play and important role in the sentiment classification.
12	The who and the what of usage of two cancer online communities [111]	To delineate patterns of use of the WebMD message board-type newsgroups related to breast and prostate cancers.	Most common communicators were survivors, but the two cancer groups - breast and prostate - on WebMD asynchronous bulletin boards, differed in percentages of communicator. Of four main categories of content - medical/treatment, intimacy/sexuality, emotional expression, and support - most common were support and medical/treatment. The groups differed significantly on those four categories, with breast cancer more support-dominated and prostate more medical/treatment-dominated and more intimacy/sexuality messages. There were no group differences in emotional-toned vs. social support seeking vs. providing support, or positive vs. negative emotions expressed. Generally, there was considerable commonality, and the differences found primarily supported gender-specific notions of communication and need.

#	Study Title	Study Purpose	Key findings
13	Can I help you? Information sharing in online discussion forums by people living with a long-term condition [112]	To explore what information is being shared on health-related discussion boards and identify the approaches people used to signpost their peers to information.	Two major themes were identified from the qualitative content analysis: 'Information Sharing from Experience' and 'Signposting Other Sources of Information'. Much of the information sharing was based on experience, which also brought in information from external sources such as health care professionals and other acknowledged experts in the field. With the rise in peer-to-peer support networks, the nature of health knowledge and expertise needs to be redefined. People online are combining external information with their own personal experiences and sharing that for others to take and develop as they wish.
14	Activity, content, contributors, and influencers of the twitter discussion on urologic oncology [113]	To analyse the activity, content, contributors, and influencers of the Twitter discussion on urologic oncology.	Twitter discussion on urologic oncology in 2014 contained 100,987 tweets created by 39,326 participants. The content analysis detected awareness, cancer, and risk as frequently mentioned words in urologic oncology tweets. Prevalently used related hashtags were the general hashtag #cancer, awareness hashtags, and the respective cancer/urology tag ontology hashtags. Health care organizations accounted for 58% of the top influencers in all cancers. The largest proportion of physicians were among the #prostatecancer and #kidneycancer (each 9%) influencers and individual contributors were most frequent in the discussion on #kidneycancer (57%) and #testicularcancer (50%). There is a significant and growing activity in the Twitter discussion on urologic oncology, particularly on #prostatecancer. The Twitter discussion is global, social, and mobile, and merits attention of stakeholders in health care as a promising communication tool.
15	Decoding twitter: Surveillance and trends for cardiac arrest and	To characterize how Twitter users seek and	Of 62,163 tweets (15,324, 25%) included resuscitation/cardiac arrest-specific information. These

#	Study Title	Study Purpose	Key findings
	resuscitation communication [114]	share information related to cardiac arrest.	tweets referenced specific cardiac arrest events (1130, 7%), CPR performance or AED use (6896, 44%), resuscitation-related education, research, or news media (7449, 48%), or specific questions about cardiac arrest/resuscitation (270, 2%). Regarding dissemination (1980, 13%) of messages were retweeted. Resuscitation specific tweets primarily occurred on weekdays. Despite a large volume of tweets, Twitter can be filtered to identify public knowledge and information seeking and sharing about cardiac arrest.
16	Tweeting About Prostate and Testicular Cancers: What Are Individuals Saying in Their Discussions About the 2013 Movember Canada Campaign? [115]	To identify the themes and messages in public conversations on Twitter about prostate and testicular cancers in response to the 2013 Movember campaign.	The major themes identified in the 2400 tweets were fundraising as a priority (34%), making a change to men's health (18%), the campaign as a moustache contest rather than a charity (26%), the use of masculine metaphors/imagery (9%), and the role of women as moustache supporters (4%). Findings from Twitter suggest that users rarely associate their campaign efforts with prostate and/or testicular cancer in public online conversations about the 2013 Movember Canada campaign.
17	Social Media and Men's Health: A Content Analysis of Twitter Conversations During the 2013 Movember Campaigns in the United States, Canada, and the United Kingdom [116]	To examine conversations that occurred in social media networking sites in order to identify whether the 2013 Movember campaign sparked conversations about prostate and testicular cancers that could potentially lead to greater awareness and understanding of the health risks men face.	Of the 12,666 tweets analysed few (n = 84, 0.7%) provided content-rich or actionable health information that would lead to awareness and understanding of men's health risks. While moustache growing and grooming was the most popular topic in U.S. tweets, conversations about community engagement were most common in Canadian and U.K. tweets. Significantly more tweets co-opted the Movember campaign to market products or contests in the United States than Canada and the United Kingdom. Findings from this content analysis of Twitter suggest that the 2013 Movember campaigns in the United States, Canada, and the United Kingdom sparked few

#	**Study Title**	**Study Purpose**	**Key findings**
			conversations about prostate and testicular cancers that could potentially lead to greater awareness and understanding of important men's health issues.
18	Temporal causality analysis of sentiment change in a cancer survivor network [117]	To uncover the causal effect (if any) of the temporally ordered posts that make up the thread on the final sentiment of the thread originators, as well as to uncover the *prima facie* causes for final sentiment in an online cancer survivor network.	The analysis of temporal causality of cancer survivor network sentiment dynamics offers new insights that the designers, managers, and moderators of an online community can utilize to facilitate and enhance the interactions so as to better meet the social support needs of participants. The proposed methodology for the analysis of temporal causality has broad applicability in a variety of settings where the dynamics of the underlying system can be modeled in terms of state variables that change in response to internal or external inputs.
19	Comparison of social support content within online communities for high-and low-survival-rate cancers [118]	To understand how emotional and informational support content differs in online communities for cancers with high and low 5-year survival rates.	High-survival-rate communities had a greater proportion of posts containing emotional support content (75%) than low-survival-rate communities (66%). Low-survival-rate communities had a greater proportion of posts containing informational support content (46%) than high-survival-rate communities (36%). Although the relationships between survival rate and support types were significant suggesting that individuals experiencing low-survival-rate cancers may have a greater desire for informational support online than individuals experiencing high-survival-rate cancers.
20	Empowering processes within prostate cancer online support groups [119]	To examine the extent to which the potentially empowering processes, previously identified in female and mixed gender online support groups, are prevalent in the naturally occurring communication in prostate cancer online support groups.	The analysis confirmed the existence of all six potentially empowering processes, including: 'information provision', 'provision of emotional support', 'provision of esteem support', 'noting similarity of experiences', 'discussion of sensitive topics' and 'chit-chat'. The discussion between members illustrated the value they attached to knowledge, mastery and control over their lives with cancer,

#	Study Title	Study Purpose	Key findings
21			strongly favouring informed choice and active participation in disease management extending knowledge of the potentially empowering processes which members may encounter as a result of participation in online support groups and suggesting that similar processes may be found in predominantly male support groups, such as prostate cancer.
21	Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups [120]	To examine possibilities for the integration of text clustering methods into the exploratory and support features of health-related applications and online interventions.	Though the clusters for each condition differed, the clusters fell into a set of common categories: Generic, Support, Patient-Centered, Experiential Knowledge, Treatments/Procedures, Medications, and Condition Management, facilitating an increased understanding of various aspects of patient experience, including significant emotional and temporal aspects of the illness experience as well as highlighting the changing nature of patients' information needs. Information provided to patients should be tailored to address their needs at various points during their illness.
22	Mining symptoms of severe mood disorders in large internet communities [121]	To further develop the categorised text corpus for depression and evaluate trained classifiers on unlabeled diabetes chat profiles.	Findings demonstrate the feasibility of scalable flagging of patients who are at risk of developing severe depression in large Internet health communities.
23	Transmissibility of the campaign for colorectal cancer awareness in Korea among Twitter users [122]	To analyze the users and the contents of Korean tweets regarding colorectal cancer and then to estimate the transmissibility of the awareness campaign among Twitter users.	Of the 10,387 tweets analysed, most were spam (n = 8,736, 84.1%), repetitively advertising unverified commercial folk remedies, followed by tweets that shared information (n = 1,304, 12.6%) and non-information (n = 347, 3.3%). Only 80.6% of the information-sharing tweets were identified as medically correct. After spam tweets had been excluded, information-sharing tweets were most frequent during the awareness campaign month.
24	Retweeting in health promotion: Analysis of tweets about Breast Cancer Awareness Month [123]	To identify the major themes contained in the tweets about Breast Cancer Awareness	Of the 1,018 #bcam tweets analysed most were about the promotional goals of fundraising or sales events compared to the

#	Study Title	Study Purpose	Key findings
25		Month and to understand the roles of retweeting in a health campaign as well as factors (e.g., content of tweets, technical characteristics of tweets, user characteristics of tweet poster) that contribute to the wide dissemination of campaign messages on Twitter.	educational goals of teaching and raising awareness about breast cancer. Tweets with photos or images exhibited greater retweetability compared to those without them, whereas the inclusion of hyperlinks or videos did not affect retweetability.
25	The wounded blogger: analysis of narratives by women with breast cancer [124]	To analyse how illness is represented (textually and graphically) in seven blogs written by Spanish women with breast cancer to determine how they reproduce a modern biomedical reasoning and a postmodern logic, offering new possibilities for empowering, generating knowledge and representing the body	Women bloggers mainly reproduce the modern logic (characterised by the restitution narrative and a predictable, disassociated and monadic body), but there are elements which break with this logic (the quest narrative, the body presented as associated, dyadic and full of desire, and the acceptance of contingency). After applying Frank's categories, it is suggested that the contemporary way of experiencing illness may question the clarity of the modern/postmodern divide.
26	Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies [125]	To evaluate the potential contribution of mining social media networks for pharmacovigilance using examples of drug–event associations that have been flagged as potential signals: rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction) and human papilloma virus (HPV) vaccine and infertility.	21% of 2,537 rosiglitazone-related posts and 84% of the 2,236 HPV vaccine-related posts referenced other web pages, mostly news items, law firms websites, or blogs. Only ten posts described personal accounts of rosiglitazone/cardiovascular adverse event experiences, and nine posts described HPV vaccine problems related to infertility. The findings suggest that publicly available data from the considered social media networks were sparse and largely untrackable for the purpose of providing early clues of safety concerns regarding the prespecified case studies.
27	A pattern-matched Twitter analysis of US cancer-patient sentiments [126]	To characterise the content of tweets authored by the United States cancer patients and use patient tweets to compute the average happiness of cancer	The most frequently tweeted cancers were breast (n = 15,421, 11% of total cancer tweets), lung (n = 2928, 2.0%), prostate (n = 1036, 0.7%), and colorectal (n = 773, 0.5%). Patient tweets pertained to the treatment course (n = 73, 26%),

#	Study Title	Study Purpose	Key findings
		patients for each cancer diagnosis.	diagnosis (n = 65, 23%), and then surgery and/or biopsy (n = 42, 15%). Computed happiness values for each cancer diagnosis revealed higher average happiness values for thyroid, breast, and lymphoma cancers and lower average happiness values for pancreatic, lung, and kidney cancers. The study confirmed that patients are expressing themselves openly on social media about their illness and that unique cancer diagnoses are correlated with varying degrees of happiness.
28	Hanging by a thread: exploring the features of nonresponse in an online young adult cancer survivorship support community [127]	To examine features of online posts that successfully solicit responses from young adult cancer survivors and assess how these differ from posts that do not solicit responses.	Analysis revealed significant differences between the sets of posts regarding content, emotions, cognitive processes, pronoun use, and linguistic complexity. More specifically, posts with replies contained fewer words per sentence, had more first-person pronouns, had more expressions of negative emotions, and contained more present tense and past tense verbs. The findings of this study can help improve peer-exchanged support in online communities so that young adult cancer survivors can more effectively receive digital support. It also provides communication researchers, health educators, and care providers a lens for understanding the YA cancer survivorship experience and helps survivors be strategic in how they use online forums to seek advice and support.
29	Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-to-consumer advertising [128]	To descriptively characterise the demographics of people using social media to discuss rheumatoid arthritis (RA) and psoriatic arthritis (PsA) and psoriasis; to evaluate the suitability of social media as a data source for drug safety research, to classify the content and timing of the posts that these	As of October 2015, there were 785,656 arthritis-related posts. Posts were predominantly US posts (75%) from patient authors (87%) under 40 years of age (61%). For herpes zoster posts (n = 815), pairwise odds ratios were significantly increased with tofacitinib versus other rheumatoid arthritis therapies. Pairwise odds ratios for mentions of perforated bowel (n=13) were higher with tocilizumab versus other therapies.

#	**Study Title**	**Study Purpose**	**Key findings**
		social media users are contributing, with a particular focus on communication related to launch of newer biologic drugs and small molecules.	
30	Predicting Asthma Prevalence by Linking Social Media Data and Traditional Surveys [129]	To extract real-time asthma-related information by mining conversations from Twitter; to link unstructured social media data with traditional structured data and to evaluate associations among asthma prevalence and public sentiment and to develop a predictive model of asthma prevalence in the US.	Asthma prevalence in the traditional survey (Behavioural Risk Factor Surveillance System, n= 464,664) using social media information collected from Twitter (n= 500,000 million tweets/day) and socioeconomic factors collected from American Community Survey (n= 3.5million per year) indicate that monitoring asthma-related tweets may provide real-time information that can be used to predict outcomes from traditional surveys.
31	“What say ye gout experts?” A content analysis of questions about gout posted on the social news website Reddit [130]	To understand patient information needs about gout by a content analysis of questions posted on Reddit.	Questions about gout posted on the subreddit site most often related to uncertainty about symptoms and disease management strategies, with infrequent questions about serum urate testing, results or targets.The single most common category of question related to uncertainty of diagnosis, with questions about disease management common. Information-seeking about medications was generally cautious, with questions about side-effects, risk of flares after starting urate-lowering therapy, and decision to start urate-lowering therapy. Community users experiencing flares posted questions about flare management, including medications, sometimes in real-time. Dietary management questions included the effectiveness of dietary changes as a management strategy, choice of alcoholic beverage, and weight loss strategies. Questions about serum urate levels were rare.
32	Organizational Twitter Use: Content Analysis of Tweets	To identify the type of communication that happens via the Twitter	A content analysis of 2916 tweets based on the Health Belief Model revealed that the content posted by

#	Study Title	Study Purpose	Key findings
	during Breast Cancer Awareness Month [131]	feeds of various organisations during Breast Cancer Awareness Month.	these organisations reflected the use of varied theoretical constructs of the framework. The study demonstrated that while different organisations shared valuable breast cancer-related content on Twitter, each used the social media platform in a different fashion, evident through focus on different types of Health Belief Model constructs while publishing breast cancer-related tweets.
33	Functional health literacy and cancer care conversations in online forums for retired persons [132]	To assess conversations within an online general health discussion boards aimed at retired persons to investigate the functional health literacy skill of older adults with cancer.	Online discussion forums hosted by retirement associations may serve as an important channel in information dissemination about cancer prevention and screening, treatment and care support and health care advice for seniors. There were 125/1817 (6.8%) Canadians and 70/892 (7.8%) US cancer discussions among participants in 2006. There were significantly more posts about provision of cancer information from the Canadian compared with the US site. American more than Canadian conversations emphasised the health system concerns as determinants of cancer care practices.
34	Psychological language on Twitter predicts county-level heart disease mortality [133]	To identify community-level psychological characteristics associated with mortality from atherosclerotic heart disease (AHD) using a data set of tens of millions of Twitter messages.	Language patterns reflecting negative social relationships, disengagement, and negative emotions-especially anger-emerged as risk factors; positive emotions and psychological engagement emerged as protective factors. A cross-sectional regression model based only on Twitter language predicted atherosclerotic heart disease mortality significantly better than did a model that combined 10 common demographic, socioeconomic, and health risk factors, including smoking, diabetes, hypertension, and obesity. Capturing community psychological characteristics through social media is feasible, and these characteristics are strong markers of cardiovascular mortality at the community level.

#	**Study Title**	**Study Purpose**	**Key findings**
35	Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions [134]	To describe an unsupervised method to generate lexicons representing the sublanguage of an online breast cancer community.	Using context vectors trained on a small seed set is a viable, robust method to expand existing medical lexicons across a range of potential semantic categories. The method was robust across semantic categories as long as seeds were good representatives of those categories. Furthermore, this study showed that the seed set can be very small (e.g., six terms like in our experiments with detecting emotion terms) and still generate viable lexicons with good coverage.
36	A qualitative evaluation of communication in Ovarian Cancer Facebook communities [135]	To investigate the content of communication in Ovarian Cancer Facebook communities in order to determine the users, to understand the content of communication and to examine the extent of exchanging incorrect health information and posting negative feeling in patient support Facebook communities.	Patients with ovarian cancer and their care takers provided queries and feedback related to personal health related information like experiences for cancer management; they also provided emotional support. Potential disadvantages such as unconventional medical information and negative feeling were infrequent.
37	Learning from social media: utilizing advanced data extraction techniques to understand barriers to breast cancer treatment [136]	To identify key issues and themes that patients with breast cancer were sharing online, focusing on barriers to treatment.	1,024,041 unique posts related to breast cancer treatment were identified of which 57% expressed negative sentiments. Treatment barriers included emotional (23% of posts), preferences and spiritual/religious beliefs (21%), physical (18%), resource (15%), healthcare perceptions (9%), treatment processes/duration (7%), and relationships (7%). Black and Hispanic (vs. white) users more frequently reported barriers related to healthcare perceptions, beliefs, and pre-diagnosis/diagnosis organizational challenges and fewer emotional barriers. Breast cancer treatment barriers differed by race/ethnicity.
38	Patient-reported outcomes as a source of evidence in off-label prescribing:	To examine the illustrative cases of amitriptyline and modafinil – drugs	There were 1948 treatment histories for modafinil and 1394 treatment reports for amitriptyline reported across five

#	Study Title	Study Purpose	Key findings
	Analysis of data from PatientsLikeMe [137]	commonly used off-label.	PatientsLikeMe communities (multiple sclerosis, Parkinson's disease, mood conditions, fibromyalgia/chronic fatigue syndrome, and amyotrophic lateral sclerosis). The majority of members reported taking the drug for off-label uses. Only 34 of the 1755 (1%) reporting purpose used modafinil for an approved purpose (narcolepsy or sleep apnea). Only 104 out of 1197 members (9%) reported taking amitriptyline for its approved indication, depression. Members taking amitriptyline for off-label purposes rated the drug as more effective than those who were taking it for its approved indication. While dry mouth is a commonly reported side effect of amitriptyline for most patients, 88 of 220 (40%) of people with amyotrophic lateral sclerosis on the drug reported taking advantage of this side effect to treat their symptom of excess saliva.
39	Type 1 Diabetes in Twitter: Who All Listen To? [138]	To describe the profile of Twitter users posting about Type 1 Diabetes and the potential relevance of the tweets, based on the number of re-tweets and followers.	Tweets posted by non-governmental organizations (NGOs), communication media, and individuals affected by T1D had higher number of potential readers. More than a half of the tweets were posted by individuals affected by T1D, and their tweets were the most re-tweeted. The next most active users were NGOs and healthcare professionals. However, while tweets soliciting for research funds posted by the NGOs were the next most re-tweeted messages, tweets posted by healthcare professionals were the least re-tweeted. Twitter could be used more actively by healthcare professionals to disseminate correct information about T1D.
40	Cancer Communication on Social Media: Examining How Cancer Caregivers Use Facebook for Cancer-Related Communication [139]	To examine how cancer caregivers use personal Facebook pages for cancer-related communication.	Six themes emerged from the analysis of 15,852 posts on 18 publically available Facebook pages hosted by parents of children with acute lymphoblastic leukemia: (1) documenting the cancer journey, (2) sharing emotional

#	Study Title	Study Purpose	Key findings
			strain associated with caregiving, (3) promoting awareness and advocacy about pediatric cancer, (4) fundraising, (5) mobilizing support, and (6) expressing gratitude for support.
41	Chronic Media Worlds: Social Media and the Problem of Pain Communication on Tumblr [140]	To explore how media networks make chronic pain matter by reimagining pain through and against biomedical frames and definitions and articulating new collective issues around the circulation of new media representations.	Temporalities reflect relations between users, technologies, and protocols, but also in the way in which users imagine health communication on social media to explore aspects of illness experience that are difficult to communicate in other contexts and to make inhabitable media worlds. Tumblr users creatively participate in framing the shortcomings of healthcare provision and services, social norms, and exclusions that affect people living with persistent, non-life-threatening chronic conditions. Health communication in social media is a response to the experience of living with pain, a way of coping with its debilitating effects, while at the same time it can be a potentially transformative act in bringing together issues and publics.
42	The thing about pain: The remaking of illness narratives in chronic pain expressions on social media [141]	To explore forms of chronic pain communication on social media platforms: Flickr and Tumblr.	Flickr brings a form of narrative immediacy, making the pain experience visible, eliciting empathy and marking chronicity. Tumblr lends itself to more networked forms of interaction through the circulation of multimodal memes, and support communities are built through humour and social criticism.
43	Dying Is Unexpectedly Positive [142]	To compare the blogs of terminally ill patients and the last words of death-row inmates with forecasts of everyday people imagining themselves facing death.	Through a comparison of two studies, it was revealed that blog posts of near-death patients with cancer and amyotrophic lateral sclerosis were more positive and less negative than the simulated blog posts of nonpatients - and also that the patients' blog posts became more positive as death neared. The second study revealed that the last words of death-row inmates were more positive and less negative than the simulated last words of noninmates - and also that these

#	Study Title	Study Purpose	Key findings
			last words were less negative than poetry written by death-row inmates. Together, these results suggest that the experience of dying even because of terminal illness or execution - may be more pleasant than one imagines.
44	Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook [143]	To qualitatively evaluate the content of communication in Facebook communities dedicated to diabetes.	Patients with diabetes, family members, and their friends use Facebook to share personal clinical information, to request disease-specific guidance and feedback, and to receive emotional support. Approximately two-thirds of posts included unsolicited sharing of diabetes management strategies, over 13% of posts provided specific feedback to information requested by other users, and almost 29% of posts featured an effort by the poster to provide emotional support to others as members of a community. 27% of posts featured some type of promotional activity, generally presented as testimonials advertising non-FDA approved, "natural" products. Clinically inaccurate recommendations were infrequent, but were usually associated with promotion of a specific product or service.
45	Automatic Extraction of Personal Experiences from Patients' Blogs: A Case Study in Chronic Obstructive Pulmonary Disease [144]	To enable easier, faster access to patient experiences shared through blog posts by facilitating content analysis using large-scale text mining.	The machine learning approach trained on lexical features successfully extracted sentences about patient experience with 93% precision and 80% recall. Automatic annotation of sentences about patient experience can facilitate subsequent content analysis by highlighting the most relevant sentences to this particular problem.
46	The invisible reality of arthritis: A qualitative analysis of an online message board [145]	To investigate how and why an online arthritis message board was used.	Eighty-seven initial messages and 981 replies were analysed identifying 4 master themes: the invisible reality of the condition, information exchange, the support offered and received online was considered to provide additional benefits to support provided by family and friends and the message board allowed users to share

#	Study Title	Study Purpose	Key findings
			(primarily negative) emotions which they felt unable to express in their offline worlds.
47	Representation of health conditions on Facebook: Content analysis and evaluation of user engagement [146]	To describe how health conditions are represented on Facebook Pages and how users interact with these different conditions.	A large number of Facebook pages (29.4%, 280/953) were irrelevant to the health condition searched. Of the 673 relevant pages, 151 were not in English or originated outside the United States, leaving 522 pages to be coded for content. The most common type of page was marketing/promotion (32.2%, 168/522) followed by information/awareness (20.7%, 108/522), Wikipedia-type pages (15.5%, 81/522), patient support (9.4%, 49/522), and general support (3.6%, 19/522). Health conditions varied greatly by the primary page type. All health conditions had some marketing/promotion pages and this made up 76% (29/38) of pages on acquired immunodeficiency syndrome (AIDS). The largest percentage of general support pages were cancer (19%, 6/32) and stomach (16%, 4/25). For patient support, stroke (67%, 4/6), lupus (33%, 10/30), breast cancer (19%, 6/31), arthritis (16%, 6/36), and diabetes (16%, 6/37) ranked the highest. Six health conditions were not represented by any type of support pages (ie, human papillomavirus, diarrhea, flu symptoms, pneumonia, spine, human immunodeficiency virus).
48	Improving knowledge of patient skills thanks to automatic analysis of online discussions [147]	To perform a systematic survey of the messages posted on French Online Discussion Forums.	Study of real exchanges between patients leads to a better understanding of their skills in daily self-management of diabetes. The method recognizes almost all the aimed skills in fora. The quality of the skills' recognition varies with the method's parameters. Most of the selected messages are relevant to at least one of the associated skills. Manual analysis shows a substantial number of messages is dedicated to daily self-care and psychosocial skills.

#	Study Title	Study Purpose	Key findings
49	Diabetes topics associated with engagement on twitter [148]	To examine engagement with diabetes information on Twitter and to examine Amazon Mechanical Turk as a new tool to aid public health researchers working with social media data.	The most common tweet topics were medical and nonmedical resources for diabetes. Tweets that included information about diabetes- related health problems were positively and significantly associated with engagement. Tweets about diabetes prevalence, nonmedical resources for diabetes, and jokes or sarcasm about diabetes were significantly negatively associated with engagement.
50	What E-patients want from the doctor-patient relationship: Content analysis of posts on discussion boards [149]	To identify what people with diabetes consider to be their role in condition management, what they consider to be their doctor’s role and what they see as positive or problematic in their interactions with medical staff.	The key themes identified in the study were ownership of condition management, power issues between people with long-term conditions and doctors, and ways in which people seek to manage their doctors. People with diabetes valued doctors who showed respect for them and their knowledge, and were willing to listen and openly discuss their options. Patients felt that they could and should take responsibility for and control of their day-to-day disease management. They saw doctors as having a role in this process, but when this was lacking, many people felt able to use alternative means to achieve their goal, although the doctor's function in terms of gatekeeping resources could create difficulties for them in this respect.
51	Health care and social media: What patients really understand [150]	To provide insight into how social media content is written at lower grade levels than traditional medical sources, improving patient health literacy.	47.6% of tweets were above ninth grade reading level. Tweets from this data sample were written at a mean grade level of 9.45, signifying a level between the ninth and tenth grades. This is higher than desired, yet still better than traditional sources, which have been previously analyzed. Ultimately, those responsible for health care social media posts must continue to improve efforts to reach the recommended reading level (between the sixth and eighth grade), so as to ensure optimal comprehension of patients.
52	Tweeting about Diabetes and Diets - Content and	To analyse the semantic content of tweets	The findings from an analysis of 9,042 tweets containing the words

#	Study Title	Study Purpose	Key findings
	Conversational Connections [151]	discussing diabetes and diets and the conversational connections of those tweeting and those being mentioned in the tweets.	"diabetes" and "diet" indicated that analysing Twitter conversations can be an efficient way to map public opinions about diabetes and diets. The results also showed that many private persons act as diabetes advocates spreading information and news about diabetes and diets. Surveying these topics can be useful for healthcare practitioners; as these are in contact with patients with diabetes, it is important that they are aware of both the most discussed topics and the most common information sources, who are often laymen.
53	Text classification for assisting moderators in online health communities [152]	To explore low-cost text classification methods to this new domain of determining whether a thread in an online health forum needs moderators' help.	Social media environments provide popular venues in which patients gain health-related information. Feature selection methods and balanced training data can improve the overall classification performance and there are implications of weighing precision versus recall for assisting moderators of online health communities. Error analysis uncovered social, legal, and ethical issues around addressing community members' needs.
54	Answers to health questions: Internet search results versus online health community responses [153]	To determine what health information search engines provide versus online health communities, how clinically accurate the information is and what the most appropriate source of health information is for different question types.	Community responses answered more questions than did search results overall. Search results were most effective in answering value questions and least effective in answering policy questions. Community responses answered questions across question types at an equivalent rate, but most answered policy questions and the least answered fact questions. Value questions were most answered by community responses, but some of these answers provided by the community were incorrect. Fact question search results were the most clinically valid.
55	Talking With death at a diner: Young women's online narratives of cancer [154]	To gain a unique perspective of young women with a diagnosis of cancer through analysis of their online illness blogs.	The narratives shared on illness blogs offer an online place for expression of emotion, information exchange, and online social support. Emotional catharsis in the young women's narrative

#	Study Title	Study Purpose	Key findings
			elucidated the experiences of transition through diagnosis, treatment, and survivorship that allow a better understanding of their emotional and psychosocial needs. Themes were identified as the women processed their diagnosis: living in the middle, new normal, urgency, and transition into the abyss.
56	Young Women's Experiences With Complementary Therapies During Cancer Described Through Illness Blogs [155]	To examine descriptions of complimentary and alternative medicine use among young women diagnosed with cancer who maintained an online cancer blog.	Women's narratives describe several themes of the experience of using complementary therapies including awakening, new identities (that incorporate loss), the good stuff, and release. Online illness blogs allow researchers to understand the complete experience of the patient through personal accounts.
57	Blogging through cancer: Young women's persistent problems shared online [156]	To describe the life disruptions caused by cancer among young women, as well as to understand the facilitators and barriers in accessing healthcare services during and after active treatment.	There were 4 dimensions of persistent problems that were articulated in the narratives of the young women without any relief. They included pain and fatigue, insurance and financial barriers, concerns related to fertility, and symptoms of posttraumatic stress and anxiety.
58	"It's Back! My Remission Is Over": Online Communication of Disease Progression Among Adolescents With Cancer [157]	To describe the experiences of adolescents with cancer who encountered disease progression through analysis of their online illness blogs.	Key themes described among the adolescents, included normalising the news, facing treatment failure, and reconciling chronos-the finite concept of time.
59	"Obesity is the New Major Cause of Cancer": Connections Between Obesity and Cancer on Facebook and Twitter [158]	To understand the way individuals discuss obesity and cancer concurrently on social media through identified relational themes, grammatical elements and valence of the sentiments in comments mentioning both conditions.	Seven relational themes emerged from analysing 1382 posts. These include obesity being associated with cancer (n = 389), additional factors being associated with both obesity and cancer (n = 335), obesity causing cancer (n = 85), cancer causing obesity (n = 6), obesity was not linked to cancer (n = 13), co-occurrence (n = 492), and obesity was valued differently than cancer (n = 60). Fifty-nine percent of posts focused on an associative or causal link between obesity and cancer. Thirty-one percent of posts contained positive and/or negative sentiment. Facebook was more

#	Study Title	Study Purpose	Key findings
			likely to contain any sentiment, but Twitter contained proportionately more negative sentiment. Concurrent qualitative analysis revealed a dominance of individual blame for overweight/obese persons and more support and empathy for cancer survivors.
60	Using Twitter to Assess the Public Response to the United States Preventive Services Task Force Guidelines on Lung Cancer Screening with Low Dose Chest CT [159]	To assess the immediate public response to the 2013 United States Preventative Services Task Force draft guidelines on lung cancer screening with Low Dose Chest CT.	Twitter activity rose rapidly after the USPSTF draft guidelines on LDCT. Most users were non-physicians and fre- quently cited non-peer-reviewed articles. Users maintained an overall favorable view of screening, while expressing various concerns. The 172 tweets in the week following release of the USPSTF guidelines on LDCT, were tweeted by 166 unique users including: news organizations/ online news gathering accounts (34.9%), general public (21.7%), physicians (12.0%, 6 radiologists), and businesses (11.4%). 23.3% of tweets provided opinion on the guidelines (50.0% favorable, 27.5% concerned toward screening). Most (91.3%) tweets contained links to a total of 46 unique articles, which were authored by lay press (41.3%), non-peer-reviewed medical press (32.6%), and hospital/medical practice websites (10.9%). Among these, 50.0% were favorable, citing mortality reduction (87.0%), published data supporting screening (50.0%), and early detection (43.5%), while 28.3% expressed concern, including false positives (58.9%) and radiation risk (39.1%).
61	The experience of young adult cancer patients described through online narratives [160]	To explore the experiences and gain a better understanding of young adults affected by cancer by examining their blogs.	The Internet provides young cancer patients with a space in which to express themselves and to share experiences with those who are of similar age and in similar situations. 10 main themes were identified: physical burdens, future prospects, isolation (physical and psychological), guilt, mortality, images of cancer, creating a positive attitude, healthcare, online

#	Study Title	Study Purpose	Key findings
			social interaction, and cancer survivorship.
62	Gender differences among young adult cancer patients: A study of blogs [161]	To determine whether blogging can provide an environment where gender norms affect young adults affected by cancer to a lesser degree than what could be expected in non-online environments.	There is little difference in blog content between genders suggesting that the blog environment could lessen the gender-typical behaviors often expected by society and may provide an outlet for young adult cancer patients to more freely share their cancer-related experiences, at the same time providing an opportunity for social connection. This is particularly significant for male patients who are known to inhibit their emotions as well as the expression of their health concerns.
63	Scientific versus experiential evidence: Discourse analysis of the chronic cerebrospinal venous insufficiency debate in a multiple sclerosis forum [162]	To understand whether scientific and experiential knowledge were experienced as contradictory in multiple sclerosis patient online forums.	There was an emotionally charged debate about CCSVI which could be generalized to 2 discourse strands: (1) the "downfall of the professional knowledge providers" and (2) the "rise of the nonprofessional treasure trove of experience." The discourse strands indicated that the discussion moved away from the question whether scientific or experiential knowledge had more evidentiary value. Rather, the question whom to trust (ie, scientists, fellow sufferers, or no one at all) was of fundamental significance. Four discourse positions could be identified by arranging them into the dimensions "trust in evidence-based knowledge," "trust in experience-based knowledge," and "subjectivity" (ie, the emotional character of contributions manifested by the use of popular rhetoric that seemed to mask a deep personal involvement).
64	Performing health identities on social media: An online observation of Facebook profiles [163]	To examine the role of Facebook in the lives of users with Type 1 and Type 2 diabetes and the multimodal discursive practices they employ in their ongoing representation of life with a long-term condition.	The analysis revealed Facebook users' sensitivity to the varied social contexts that were collapsed within their networks as well as the strategies they employed to perform publically acceptable identities. Salient multimodal actions performed by participants included constructing personal expertise in relation to diabetes

#	**Study Title**	**Study Purpose**	**Key findings**
			management, displaying the individual's integration into wider diabetes-related networks, presenting mundane aspects of self-management verbally and visually, and adopting a playful stance.
65	Tweeting influenza vaccine to cardiovascular health community [164]	An assessment of patients' perception of the influenza vaccine and the reason for its underutilisation, by mining data from Twitter.	The content of the 29,243 tweets associated with the influenza vaccine, was characterized based on tone, emotion, awareness, perception, style, and perspective regarding the influenza vaccine. The tweets often pertained to self-reports after receiving the influenza vaccine (14%); the reason for not receiving the influenza vaccine (12%); emotional language with positive or negative sentiments (33%); and advertisement, news, or updated research (41%). For example, many tweets acknowledged the reasons for receiving or not receiving the influenza vaccine in individuals with cardiovascular disease.
66	Real-Time disease surveillance using twitter data: Demonstration on flu and cancer [165]	To describe a novel real-time flu and cancer surveillance system that uses spatial, temporal and text mining on Twitter data.	The US disease surveillance maps, reveal distribution and timelines of disease types, symptoms, and treatments, in addition to overall disease activity timelines. This surveillance system can be very useful not only for early prediction of seasonal disease out- breaks but also for monitoring distribution of cancer patients with different cancer types and symptoms in each state and the popularity of treatments used.
67	Raising Awareness About Cervical Cancer Using Twitter: Content Analysis of the 2015 #SmearForSmear Campaign [166]	To identify the tweets delivering a raising awareness message promoting cervical cancer screening and the characteristics of Twitter users posting about this campaign.	This study demonstrated that the success of the Twitter campaign called #SmearForSmear launched in 2015 for the European Cervical Cancer Prevention Week. A total of 1881 tweets were analysed. 70% of tweets had been posted by people living in the United Kingdom. A total of 57% of users were women, and sex was unknown in 36% of cases. 54% of the users had posted at least one selfie with smeared lipstick. Independent factors associated with posting sensitising tweets were

#	Study Title	Study Purpose	Key findings
			women who experienced an abnormal smear test, female gender, and people who live in the United Kingdom. Non-sensitising tweets were statistically more posted by a nonhealth or nonmedia company.
68	Identifying adverse drug events from patient social media: A case study for diabetes [167]	To describe a high-performance automatic information extraction process for identifying adverse drug events on social media.	Compared to a fully manual approach, the proposed method minimized manual effort and managed to improve efficiency. Compared to baseline methods, the approach significantly improved the accuracy and overall quality of the social media adverse drug event reports, which provides more reliable evidence for risky drug identification.
69	A Qualitative Analysis of "naturalistic" Conversations in a Peer-Led Online Support Community for Lung Cancer [168]	To describe the content of messages in a United States-based online support community for lung cancer.	Most users were female and equally divided as patients or support persons. Content analysis generated 9 themes: disease information, diagnostic test information, treatment information, symptoms, marked deterioration, advocacy, experiencing healthcare providers and the system, positive survivorship, and making sense of emotions. Findings highlighted how the online support community is a valued, accessible avenue for information exchange and nonjudgmental emotional support for individuals dealing with lung cancer.
70	Tweet this: how advocacy for breast and prostate cancers stacks up on social media [169]	To perform a comprehensive comparison of Twitter activity related to prostate and breast cancer over the past five years.	Twitter activity has increased substantially over time for both breast and prostate cancers. Breast cancer-related activity has always been substantially higher, even during prostate cancer awareness months. The breast cancer Twitter community involved more patients and celebrities, whose disease disclosures have been previously shown to boost online activity. By contrast, physicians played a larger role in prostate cancer discussions on Twitter. The key influencers also differed between #BreastCancer and #ProstateCancer compared to #BCSM and #PCSM. Who is

#	**Study Title**	**Study Purpose**	**Key findings**
			tweeting about a topic can influence the nature and value of the discussion.
71	Work and "mass personal" communication as means of navigating nutrition and exercise concerns in an online cancer community [170]	Using the theory of illness trajectories as a framework to explore how work associated with living with a chronic illness, such as cancer, impacts and is impacted by nutrition and exercise concerns.	An analysis of discussion board threads in an online community for young adults dealing with cancer showed that nutrition and exercise needs affected the young adults' illness trajectories, including their management of illness, everyday life, biographical, and communication work.
72	Health-Related Hot Topic Detection in Online Communities Using Text Clustering [171]	To automatically distinguish different health-related topics in online health communities more effectively using text clustering.	Experiment results demonstrate that health-related hot topics primarily include symptoms, examinations, drugs, procedures and complications. Further analysis reveals that there also exist some significant differences among the hot topics discussed on different types of disease discussion boards
73	Understanding Health Care Social Media Use From Different Stakeholder Perspectives: A Content Analysis of an Online Health Community [172]	To develop a new content analysis method using text-mining techniques applied in health care social media to identify different health care stakeholders, to determine hot topics of concern and to measure sentiment expression by different stakeholders.	Three significantly different stakeholder groups were identified: patients (24,429/39,606, 61.68%) and caregivers (12,232/39,606, 30.88%) represented the majority of the population, in contrast to specialists (2945/39,606, 7.43%). Five significantly different health-related topics were identified: symptom, examination, drug, procedure, and complication. Patients were concerned most about symptom topics related to lung cancer, drug topics related to diabetes, and examination topics related to breast cancer. By comparison, caregivers were more concerned about drug topics related to lung cancer, procedure topics related to breast cancer, and complication topics. In addition, patients were more likely than caregivers and specialists to express their emotions. However, patients' sentiment intensity score was lower than those of caregivers and specialists. In particular, for caregivers, negative sentiment scores were higher than positive scores, with the opposite among specialists. Overall, the proportion of negative messages was greater

#	**Study Title**	**Study Purpose**	**Key findings**
74			than that of positive messages related to symptom, complication, and examination. The pattern was opposite for drug and procedure topics.
74	"5 Mins of uncomfyness is better than dealing with cancer 4 a lifetime": An exploratory qualitative analysis of cervical and breast cancer screening dialogue on twitter [173]	To ascertain if cancer screening content occurs on the Twitter platform, if there is sufficient volume for meaningful analysis and to understand the nature of the content.	Over three fourths of the 203 Pap smear and 271 mammogram messages were from individual accounts. Overall, 22% of Pap smear messages and 25% of mammogram messages discussed personal experiences, including attending appointments, negative sentiment about the procedure, and results. Other messages from both individuals and organisations (8% Pap smear, 18% mammogram) promoted screening. About one quarter of the messages expressed personal experiences with cancer screening. Twitter can be a rich source of information and could be used to design new health-related interventions.
75	Online discussion of drug side effects and discontinuation among breast cancer survivors [174]	To understand the frequency and content of side effects and associated adherence behaviours discussed by breast cancer patients related to using aromatase inhibitors (AIs), with particular emphasis on AI-related arthralgia.	Among 25,256 posts related to aromatase inhibitors, 4,589 (18.2%) mentioned at least one side effect. Top-cited side effects on message boards related to aromatase inhibitors were joint/musculoskeletal pain (N=5,093), hot flashes (1,498), osteoporosis (719), and weight gain (429). Among the authors posting messages who self-reported aromatase inhibitors use, 12.8% mentioned discontinuing aromatase inhibitors, while another 28.1% mentioned switching aromatase inhibitors. Although patients often cited severe joint pain as the reason for discontinuing aromatase inhibitors, many also offered support and advice for coping with aromatase inhibitors-associated arthralgia.
76	Increase in contralateral prophylactic mastectomy conversation online unrelated to decision-making [175]	To identify if the rate of conversation about contralateral prophylactic mastectomies on an online health community is increasing and what	Surgically related posts (e.g., mentioning at least one operation) made up a small percentage (n = 27,678; 8.4%) of all posts on this community. The percentage of surgically related posts mentioning CPM was found to increase over

#	Study Title	Study Purpose	Key findings
		factors may be contributing to this trend.	time, whereas the percentage of surgically related posts mentioning mastectomy decreased over time. Among posts that mentioned more than one operation, mastectomy and lumpectomy were the procedures most commonly mentioned together, followed by mastectomy and CPM. There was no change over time in the frequency of posts that mentioned more than one operation. Mosts posts mentioning a single operation were unrelated to surgical decision-making; rather the operation was mentioned only in the context of the patient's cancer history. Conversely, the most posts mentioning multiple operations centered around the patients' surgical decision-making process.
77	Symptom clusters in women with breast cancer: an analysis of data from social media and a research study [176]	To compare and contrast symptom cluster patterns derived from messages on a breast cancer forum with those from a symptom checklist completed by breast cancer survivors participating in a research study.	The following clusters were identified using forum data: menopausal/psychological, pain/fatigue, gastrointestinal, and miscellaneous. Study data generated the clusters: menopausal, pain, fatigue/sleep/gastrointestinal, psychological, and increased weight/appetite. Although the clusters are somewhat different, many symptoms that clustered together in the social media analysis remained together in the analysis of the study participants. Density of connections between symptoms, as reflected by rates of co-occurrence and similarity, was higher in the study data.
78	Patient Understanding of the Risks and Benefits of Biologic Therapies in Inflammatory Bowel Disease: Insights from a Large-scale Analysis of Social Media Platforms [177]	To use social media data to examine Inflammatory Bowel Disease patients’ understanding of the risks and benefits associated with biologic therapies and how this affected decision-making in IBD treatment.	452 (28.3%) of the 1,598 IBD-related posts, centered on the risks and/or benefits of biologics. There were 5 main themes: negative experiences and concerns with biologics (n = 247; 54.6%), decision-making surrounding biologic use (n = 169; 37.4%), positive experiences with biologics (n = 168; 37.2%), information seeking from peers (n = 125; 27.7%), and cost (n = 38; 8.4%). Posts describing negative experiences primarily commented

#	Study Title	Study Purpose	Key findings
			on side effects from biologics, concerns about potential side effects and increased cancer risk, and pregnancy safety concerns. Posts on decision-making focused on nonbiologic treatment options, hesitation to initiate biologics, and concerns about changing or discontinuing regimens.
79	LIVING WITH CANCER: Affective Labour, Self-expression And The Utility of Blogs [178]	To understand the kinds of personal investment or labour involved in the process of forming and maintaining cancer blogs over a sustained period.	The value of cancer blogging is personal, in the form of identity and affect management, network-enabling in generating online spaces for shared experience and support, and social in what is recouped in the forms of non-institutional management of serious illness. In addition, blogging helps to shape the broader social understanding of cancer, its experience and personal affects.
80	Towards Automatic Pharmacovigilance: Analysing Patient Reviews and Sentiment on Oncological Drugs [179]	To utilise pharmaceutical drugs reviews by patients on various health communities to identify frequently occurring issues and to compare these issues with food and drug administration (FDA) approved drug labels for possible improvements.	A robust and scalable system was demonstrated for automatic pharmacovigilance by extracting information from online health forums to estimate the performance of the drugs and to continuously collect, detect and monitor for occurrence of new adverse effects, once the license for the drugs has been approved by regulatory bodies. The creation of rich corpuses for certain identified sections of the drug labels and development of a mapping for oncological indications and interventions assisted the system to suggest common, rare and recommended term for different sections of the approved drug labels. Findings based on sentiment analysis of the user comments and generation of rating for the drugs shows the user perceptive regarding the drugs.
81	Social Networking Sites for Peer-Driven Health Communication: Diabetes-Related Communities in Google+ [180]	To characterise the Google+ communities related to diabetes and identify the factors associated with the activity level of these communities.	145 Google+ communities and 378 posts were included in the analysis. Majority (80.0%) of the communities were focused on “General Diabetics”, 11.8% were focused on “Type 1 diabetes”, 6.2% were focused on “Type 2 diabetes”, and 2.1% were focused

#	Study Title	Study Purpose	Key findings
			on “Gestational diabetes”. Majority of the communities (35.2%) had the principal objective “Provide support and share experience”. Majority of the posts (69.6%) were focused on “General diabetes”, 16.9% were focused on “Type 1 diabetes”, 12.4% were focused on “Type 2 diabetes” and 1.15 were focused on “gestational diabetes”. The top two themes of the posts were “Providing information” (72.8%) and “Advertisement” (31.5%).
82	#Stupidcancer: Exploring a Typology of Social Support and the Role of Emotional Expression in a Social Media Community [181]	To explore the intersections of social support, emotional expression and message sharing in an online grass-roots cancer community.	Two-thirds (66.6%) of all tweets in the sample expressed some form of informational support (n = 1,304). The next most common form of social support was information reception, present in 12.9% of tweets (n = 252). Emotional expression, present in 12.0% of tweets, was nearly as common as information reception (n = 234), while emo- tional reception was the least frequent type of social support found in the sample (4.6%, n = 91). Hope was positively related to information giving, encouragement, detailed explanations, giving thanks, and religion, and negatively related to link giving. Sadness was positively related to detailed explanation, religion, and empathy, and negatively related to information giving, giving thanks, and link giving. Fear was positively related to information seeking, detailed explanations, and empathy, and negatively related to link presence. Humour was positively related to giving thanks and detailed explanations, and negatively related to link presence. Anger was positively related to detailed explanation and negatively related to link presence.
83	Putting the pieces together: Endometriosis blogs, cognitive authority, and collaborative information behavior [182]	To understand how bloggers present information sources and make cases for and against the authority of	Bloggers invoked many sources of evidence, including experiential, peer-provided, biomedical, and intuitive ones, when discussing and refuting authority. Additionally,

#	Study Title	Study Purpose	Key findings
		those sources in peer-written blogs about chronic illness endometriosis.	they made and disputed claims of cognitive authority via two interpretive repertoires: a concern about the role and interests of the pharmaceutical industry and an understanding of endometriosis as extremely idiosyncratic. Affective authority of information sources was also identified, which presented as social context, situational similarity, or aesthetic or spiritual factors.
84	Existential challenges in young people living with a cancer diagnosis [183]	To elucidate if the theoretical framework of Yalom and his four 'givens' were expressed in blogs written by young adults living with various cancer diagnoses in Sweden.	Young adults are empowered by the writing of blogs and blogs can play an important part in increasing wellbeing and a sense of coherence within this population. The findings offered are described as a journey with several existential challenges and with death as an impending threat. The bloggers' awareness of their mortality was described as creating a sense of loss and existential loneliness.
85	Cancer information-seeking behaviors and information needs among Korean Americans in the online community [184]	To identify cancer information seeking behaviour and information needs among Korean Americans.	Medical topics accounted for 71.4 % of the main topics of the 273 postings. Treatment was the most frequently discussed in the medical topics while in the non-medical category, the most frequently discussed topic was recommendations for hospitals or doctors. In relation to types of cancer, breast cancer was the greatest concern, followed by cervical and liver cancer.
86	The source and credibility of colorectal cancer information on twitter [185]	To investigate the information sources and attempt to evaluate the credibility of colorectal cancer information in tweet content.	Individual users authored the majority of the 76,119 tweets related to colorectal cancer (n=68,982, 90.6%). They mostly tweeted about news articles/research (n=16,761, 22.0%) and risk/prevention (n=14,767, 19.4%). Medical professional users generated only 2.0% of total tweets (n=1509), and medical institutions rarely tweeted (n=417, 0.6%). Organizations tended to tweet more about information than did individuals. Credibility analysis of medically relevant sample tweets revealed that most were medically correct (n=1763, 84.5%). Among

#	**Study Title**	**Study Purpose**	**Key findings**
			those, more frequently retweeted tweets contained more medically correct information than randomly selected tweets.
87	The transition of stroke survivors from hospital to home: Understanding work and design opportunities [186]	To develop a richer understanding of the work of stroke survivors and their caregivers in the transition from hospital to home.	The thematic analysis of 318 posts highlighted the breadth of work on top of rehabilitation and other efforts to manage the impact of a stroke. The outcome of a stroke can affect everyday life work, from household chores to transportation to managing relationships. Stroke survivors also engage in biographical work in an attempt to reconstruct their personal narrative.
88	Who's talking about breast cancer? Analysis of daily breast cancer posts on the internet [187]	To assess the number, content and accuracy of new internet postings relating to breast cancer posted online worldwide in the English language over a one-month period and to compare the number of breast cancer-related postings with postings relating to colon, rectal, lung and prostate cancer.	Breast related topics had significantly more posts per day compared to others. Most posts were on media sites (65.8%). Accuracy levels were high (87.5%) but significantly lower where posted on blogs and discussion forums. Anonymous posts were common (55%) and less likely to be accurate. Use of discussion forums has exponentially increased over the last five years.
89	Unleashing the true potential of social networks: confirming infliximab medical trials through Facebook posts [188]	To provide insight into how a growing community of chronic patients, Crohn’s disease patients, employ online tools.	The quantitative analysis identified Infliximab as the most controversial argument of debate on Facebook, among Crohn's disease patients, and a second qualitative one which confirms such finding, while also comparing it to the medical literature available on such subject. The ethnographic comparative analysis which was carried out with the medical literature suggest that controversial outcomes are registered by Crohn’s disease patients using Infliximab, which can only be understood when the duration of the treatment, and the size of the population involved in the trial are taken into consideration.
90	Health Communication in Social Media: Message Features Predicting User Engagement on Diabetes-Related Facebook Pages [189]	To analyse health communications within ten diabetes-related Facebook pages to identify message	Multi-level, negative binomial regressions revealed that specific features predicted different forms of engagement. Imagery emerged as a strong predictor; messages with images had higher rates of

#	Study Title	Study Purpose	Key findings
		features predictive of user engagement.	liking and sharing relative to messages without images. Diabetes consequence information and positive identity predicted higher sharing while negative affect, social support, and crowdsourcing predicted higher commenting. Negative affect, crowdsourcing, and use of external links predicted lower sharing while positive identity predicted lower commenting. The presence of imagery weakened or reversed the positive relationships of several message features with engagement. Diabetes control information and negative affect predicted more likes in text-only messages, but fewer likes when these messages included illustrative imagery. Similar patterns of imagery's attenuating effects emerged for the positive relationships of consequence information, control information, and positive identity with shares and for positive relationships of negative affect and social support with comments.
91	Patients' perspectives on self-testing of oral anticoagulation therapy: content analysis of patients' internet blogs [190]	To elicit patients' perspectives and experiences regarding prothrombin time/international normalised ratio self-testing using portable coagulometer devices.	The key themes in relation to self-testing of oral anticoagulation therapy (OAT) identified were as follows: patient benefits reported were time saved, personal control, choice, travel reduction, cheaper testing, and peace of mind. Equipment issues included high costs, reliability, quality, and learning how to use the device. Prothrombin time (PT) and the international normalised ratio (INR) issues focused on the frequency of testing, INR fluctuations and individual target (therapeutic) INR level. Other themes noted were INR testing at laboratories, the interactions with healthcare professionals in managing and testing OAT and insurance companies' involvement in acquiring the self-testing equipment. Social issues included the pain and stress of taking and testing for OAT.

#	Study Title	Study Purpose	Key findings
92	Giving and receiving peer advice in an online breast cancer support group [191]	To examine the ways in which peers exchange advice within an online breast cancer forum, in order to better understand online groups as a resource for decision-making.	The majority of the messages solicited advice through problem disclosure or requests for information and opinion. Women construct their advice requests to target like-minded people.
93	Communicating health decisions: An analysis of messages posted to online prostate cancer forums [192]	To examine the kinds of treatment choices prostate cancer patients are making and the reports of their decision-making processes to peers through an online environment in the context of the Heuristic Systematic Model.	Men looking online for information about treatment options for Prostate Cancer are exposed to a range of decision-making processes. 49.6% of the 137 messages selected from blogs and online forums reported non-systematic decision processes, with deferral to the doctor and proof of cancer removal being the most common. For systematic processing (36.5%), messages most commonly considered treatment outcomes and side-effects. Processes did not vary between the blogs and online forums.
94	Twitter as a potential data source for cardiovascular disease research [193]	To characterise the volume and content of tweets related to cardiovascular disease and the characteristics of Twitter users.	Of the 550,338 tweets associated with cardiovascular disease, the terms diabetes (n = 239,989) and myocardial infarction (n = 269,907) were used more frequently than heart failure (n = 9414). Users who tweeted about cardiovascular disease were more likely to be older than the general population of Twitter users and less likely to be male. Most tweets were associated with a health topic; common themes of tweets included risk factors, awareness, and management of cardiovascular disease.
95	Identifying Consumer Value Co-created through Social Support within Online Health Communities [194]	To provide insight into how consumers experience co-created value within online health communities through social support.	Online Parkinson's Disease and ALS communities give consumers the opportunity to create and experience forms of consumer value that would not otherwise be available in a traditional health delivery system.
96	Beyond Words: Amplification of Cancer Risk Communication on Social Media [195]	To examine the role of audiences in the process of mediated risk communication with messages posted on the National Cancer	This study showed that the social amplification of risk framework, despite being developed before social networking sites existed, provides a robust theoretical framework for the analysis of message dissemination and social

#	Study Title	Study Purpose	Key findings
		Institute's Facebook page.	amplification of risk information. This study confirmed that risk messages are amplified through the means of social communication channels, that information amplification occurs throughout the different levels of communication behaviour and can be observed among the audiences who communicate actively through comments, who amplify messages to their circles through sharing, and who passively participated in the communication process by publicly liking the messages. Social amplification does not happen through the repetition of risk-related keywords as communication signals.
97	A Picture Really is Worth a Thousand Words: Public Engagement with the National Cancer Institute on Social Media [196]	To assess the role of information richness of the messages disseminated by the National Cancer Institute through Facebook and the relative effectiveness of these social media messages to promote engagement with cancer prevention and research information.	Audience engagement is associated with the format of cancer-related social media posts. Specifically, posts with photos received significantly more likes, comments, and shares than videos, links, and status updates. This has implications for how social media can be more effectively utilised to promote public engagement with important public health issues.
98	Comment topic evolution on a cancer institution's facebook page [197]	To gain a more accurate understanding of oncology patients' collective care experiences by automatically extracting topics from comments posted by consumers on a cancer institutions' social media page, to see how the quantity of topics evolved over time and to identify significant correlations between topics.	Ten topics were identified in the 4,335 comments posted on Facebook: greetings (17.3%), comments about the cancer institution (16.7%), blessings (10.9%), time (10.7%), treatment (9.3%), expressions of optimism (7.9%), tumor (7.5%), father figure (6.3%), and other family members & friends (8.2%), leaving 5.1% of comments unclassified. There was a strong positive correlation between greetings and other family members & friends, a positive correlation between blessings and the cancer institution, and a negative correlation between blessings and greetings.
99	Using Twitter for breast cancer prevention: An	To understand the use of Twitter during Breast Cancer Awareness	The number of tweets spiked dramatically during the first few days of breast cancer awareness

#	**Study Title**	**Study Purpose**	**Key findings**
	analysis of breast cancer awareness month [198]	Month by analysing the frequency of tweeting about breast cancer; the individuals, organisations, or celebrities more likely to tweet; and the reach and content of these messages.	month and then tapered off. There was an average of 1.69 tweets per user. The majority of the 797,827 users were individuals. Nearly all of the 1,351,823 tweets were original. Organisations and celebrities posted more often than individuals. On average celebrities made far more impressions; they were also retweeted more often and their tweets were more likely to include mentions. Individuals were more likely to direct a tweet to a specific person. Organisations and celebrities emphasized fundraisers, early detection, and diagnoses while individuals tweeted about wearing pink.
100	Do cancer patients tweet? Examining the twitter use of cancer patients in Japan [199]	To understand cancer patients' social media usage and gain insight into patient needs.	This study demonstrates that cancer patients share information about their underlying disease, including diagnosis, symptoms, and treatments, via Twitter.Twitter user profiles included references to breast cancer (n=313), leukemia (n=158), uterine or cervical cancer (n=134), lung cancer (n=87), colon cancer (n=64), and stomach cancer (n=44). A co-occurrence network was seen for all of these cancers, and each cancer had a unique network conformation. Keywords included words about diagnosis, symptoms, and treatments for almost all cancers. Words related to social activities were extracted for breast cancer. Words related to vaccination and support from public insurance were extracted for uterine or cervical cancer.
101	Patient perspectives of dabigatran: Analysis of online discussion forums [200]	To determine patient experiences and perceptions regarding dabigatran through qualitative thematic content analysis of comments posted on publicly accessible virtual discussion forums and Internet support groups.	Five broad thematic categories emerged from the posted comments: general concerns about safety and efficacy, questions about indications and contraindications, questions about proper use and storage, questions about diet and drug restrictions, and experiences with perceived side effects. A primary concern for patients taking dabigatran was the lack of antidote to reverse the effects of dabigatran if bleeding occurs. Several

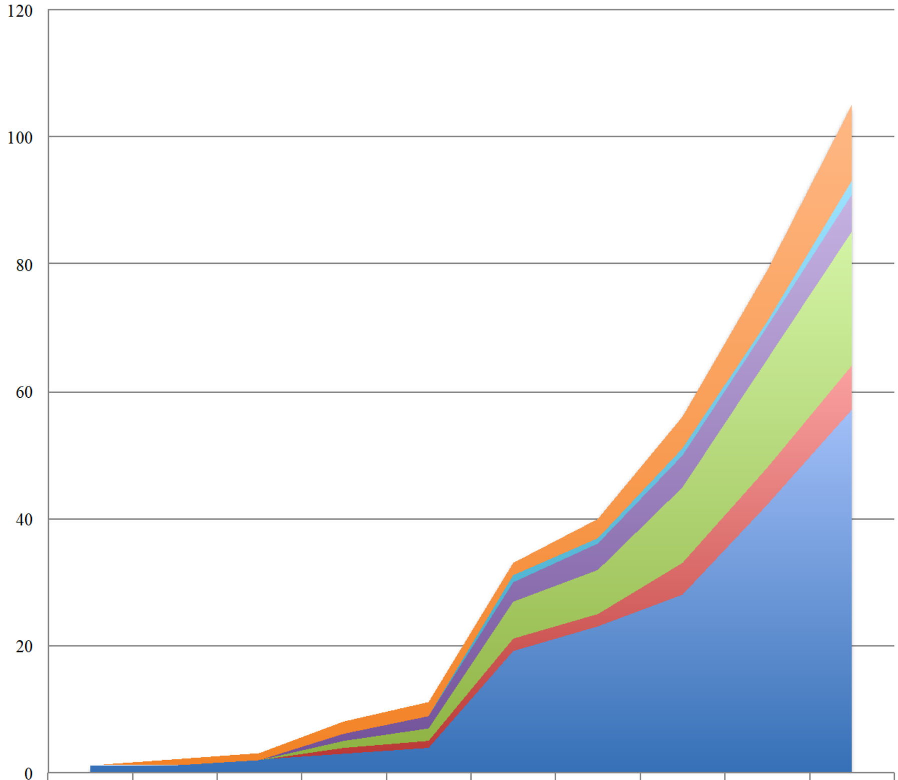
#	Study Title	Study Purpose	Key findings
102			questions pertaining to the use of dabigatran with other medications or medical conditions were noted, and multiple patients expressed confusion about instructions for using dabigatran before and after medical procedures. An unexpected finding included several criticisms of the medication packaging, which many patients found inconvenient or difficult to open. Several perceived side effects were noted, including some not reported in clinical trials.
102	A scalable framework to detect personal health mentions on Twitter [201]	To develop a scalable framework to detect personal health status mentions on Twitter and assess the extent to which such information is disclosed.	Tweets from a small subset of the health issues can train a scalable classifier to detect health mentions.Training on 2000 tweets from four health issues (cancer, depression, hypertension, and leukemia) yielded a classifier with precision of 0.77 on all 34 health issues. Twitter users disclosed personal health status for all health issues. Personal health status was disclosed over 50% of the time for 11 out of 34 (33%) investigated health issues. The disclosure rate was dependent on the health issue in a statistically significant manner. For instance, more than 80% of the tweets about migraines (83/100) and allergies (85/100) communicated personal health status, while only around 10% of the tweets about obesity (13/100) and heart attack (12/100) did so. The likelihood that people disclose their own versus other people's health status was dependent on health issue in a statistically significant manner. For example, 69% (69/100) of the insomnia tweets disclosed the author's status, while only 1% (1/100) disclosed another person's status. By contrast, 1% (1/100) of the Down syndrome tweets disclosed the author's status, while 21% (21/100) disclosed another person's status. It is possible to automatically detect personal health status mentions on Twitter in a scalable manner.

#	Study Title	Study Purpose	Key findings
103	Talking About My Care: Detecting Mentions of Hormonal Therapy Adherence Behavior in an Online Breast Cancer Community [202]	To develop a machine-learning framework to distinguish mentions of hormonal therapy adherence behaviour from other less relevant free-text contents in online health forums.	Shallow neural networks, in the form of wordlvec, are show to have learned features that can be applied to build efficient hormonal therapy adherence behavior mention classifiers. Through medical term comparison, it was found that patients who exhibit an interruption behaviour were more likely to mention depression and their care providers, while patients with continuation behavior were more likely to mention common side effects (e.g., hot flashes, nausea and osteoporosis), vitamins and exercise.
104	Towards Supporting Patient Decision-making In Online Diabetes Communities [203]	To better understand the information needs of patients with diabetes and how this pertains to their decision-making.	The analysis of 801 posts from an online diabetes community revealed motivations for posters' inquiries related to decision-making including the changes in disease state, increased self-awareness, and conflict of information received. Medication and food were the among the most popular topics discussed as part of their decision-making inquiries.
105	Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks [204]	To provide an annotation schema for topic classification; to contribute an annotated dataset of sentences and posts according to the coding schema; to experiment with different supervised classification tools; and to explore the prevalence and dynamics of different discussion topics in the entire breast cancer community and across member with different disease severities.	The results suggested that convolutional neural network outperforms other classifiers in the task of topic classification and identify several patterns and trajectories in a popular breast cancer online community. For example, although members discuss mainly disease-related topics, their interest may change through time and vary with their disease severities.

The type of NCDs investigated in the existing literature

The cumulative volume, over time, of the type of NCDs investigated in the 105 eligible studies is described in Figure 8, where cancer and diabetes are seen to dominate the existing literature at 54% (n=57) and 20% (n=21) respectively.

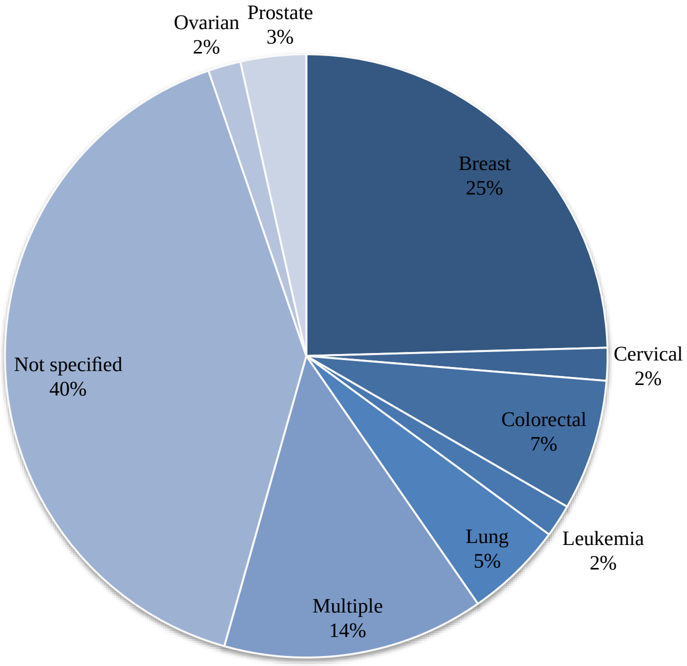
Figure 8. Cumulative volume, between 2007-2017, of eligible studies by type of NCD investigated



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
🟧 Other chronic conditions	0	1	1	2	2	2	3	5	8	12
🟦 Respiratory	0	0	0	0	0	1	1	1	1	2
🟪 Multiple	0	0	0	1	2	3	4	5	5	6
🟩 Diabetes	0	0	0	1	2	6	7	12	17	21
🟥 Cardiovascular	0	0	0	1	1	2	2	5	6	7
🟦 Cancer	1	1	2	3	4	19	23	28	42	57

When looking with a higher level of granularity into the type of cancers being investigated (Figure 9), non-specified cancers (40%, n=23), breast cancer (25%, n=14) and studies that include multiple cancer types (14%, n=8) dominate the existing literature.

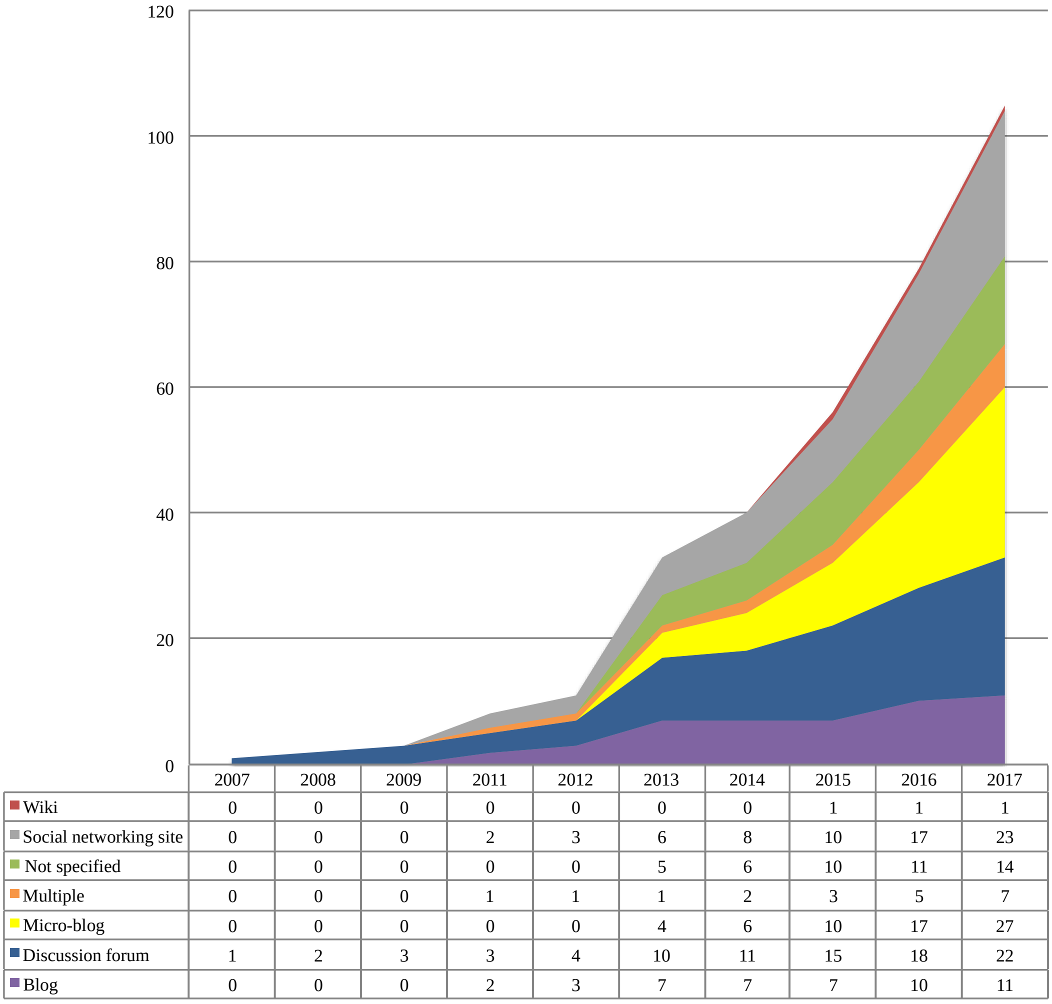
Figure 9. Percentage of the types of cancers being investigated in the 105 eligible NCD studies



Social media platform used as source of data

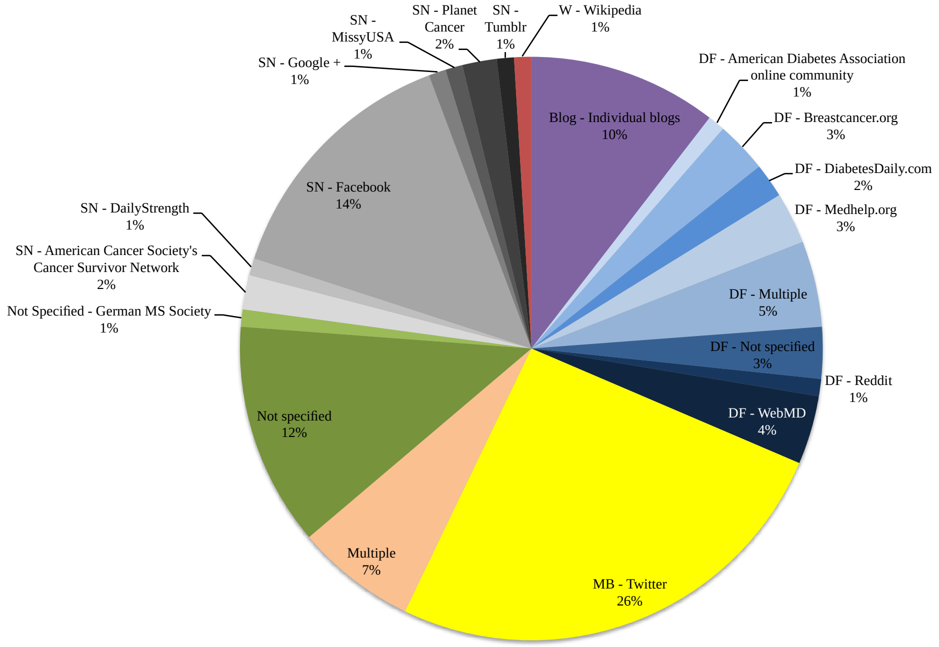
The cumulative volume, over time, of the types of social media [5] from which data were sourced for the eligible NCD studies is described in Figure 10, where micro-blogging sites (i.e. Twitter), social networking sites and discussion forums are predominately used as sources of data for 26% (n=27), 22% (n=23) and 21% (n=22) of the eligible NCD studies.

Figure 10. Cumulative volume, between 2007-2017, of the eligible NCD studies classified by type of social media from which the data was sourced



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
🟥 Wiki	0	0	0	0	0	0	0	1	1	1
■ Social networking site	0	0	0	2	3	6	8	10	17	23
🟩 Not specified	0	0	0	0	0	5	6	10	11	14
🟧 Multiple	0	0	0	1	1	1	2	3	5	7
🟨 Micro-blog	0	0	0	0	0	4	6	10	17	27
🟦 Discussion forum	1	2	3	3	4	10	11	15	18	22
🟪 Blog	0	0	0	2	3	7	7	7	10	11

Figure 11. Percentage of named social media from which the data in the eligible NCD studies were sourced



Key: DF = Discussion forum, MB = Micro-blog, SN = Social networking site, W = Wiki

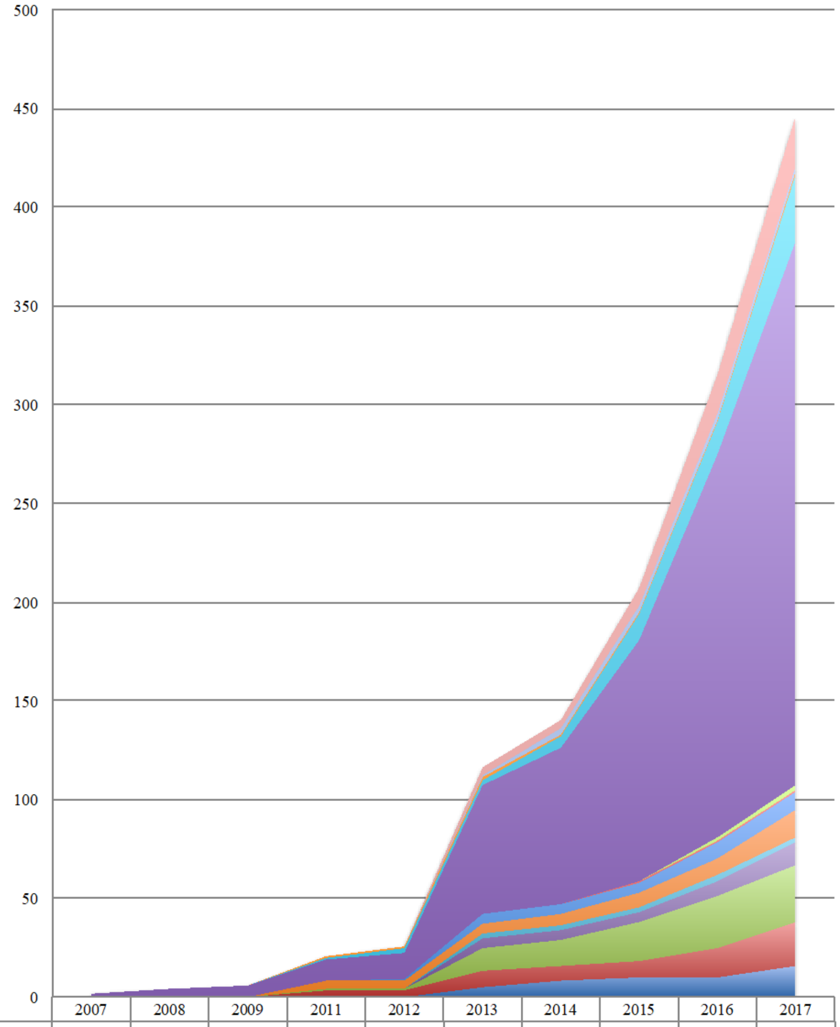
By breaking this down by named social media (Figure 11), it is evident that mainstream social media such as Twitter (26%, n=27), Facebook (14%, n=15) and individual blogs (10%, n=11) were the main sources of data for the eligible studies, compared to Wikipedia (1%, n=1), Google+ (1%, n=1), Tumblr (1%, n=1) and health-specific social media such as the discussion forums on DiabetesDaily.com (2%, n=2), breastcancer.org (3%, n=3), medhelp.org (3%, n=3) and webMD.org (4%, n=4).

Academic communities most active in this field of research

Given the multi-disciplinary nature of this research, the subject areas [96] of the author affiliation, the sources of funding for conducting this type of research, as well as the journals in which the studies were published, were analysed.

The cumulative volume, over time, of the subject areas to which all the authors were affiliated is described in Figure 12. Medicine is the dominant subject area, with 63% (n=276) of the 435 unique authors affiliated to this discipline, followed by computer sciences (7%, n=29), technology (6%, n=25) and communication (5%, n=22). Eight per cent (n=34) of the authors did not specify which discipline they were affiliated to, whilst 3% (n=14) of the authors of the eligible studies were affiliated to non-academic or research institutions such as Google, Treato, Bristol Myers Squibb and Takeda.

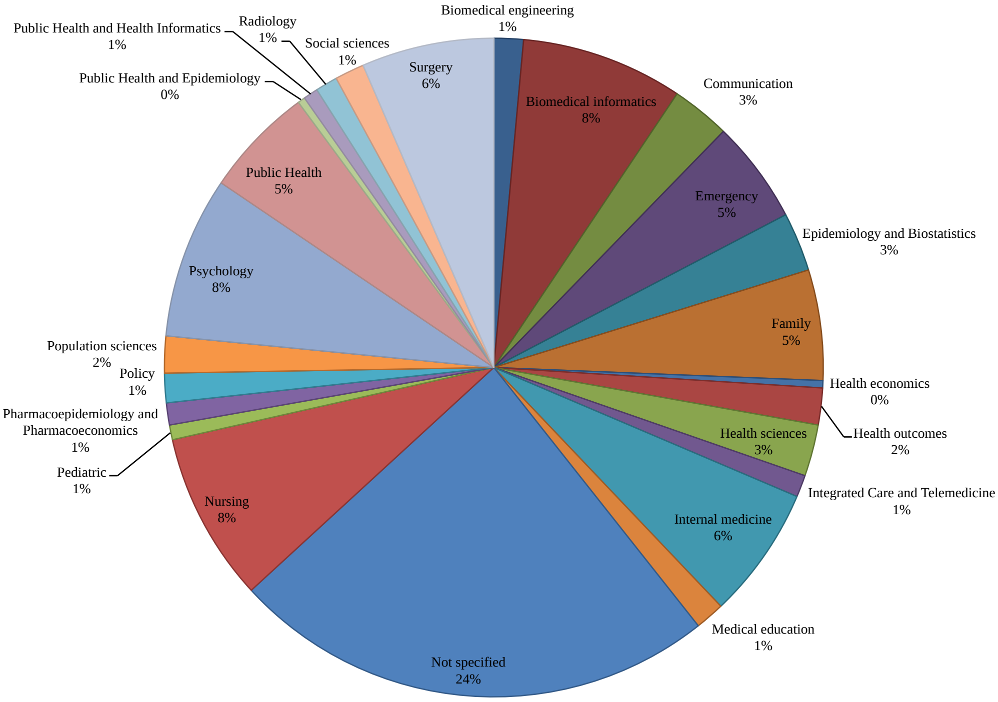
Figure 12. Cumulative volume between 2007-2017 of eligible NCD studies by subject area of authors' affiliation



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
Technology	0	0	0	0	0	4	4	9	20	25
Science and Engineering	0	0	0	0	0	1	3	3	3	3
Politics	0	0	0	1	1	1	1	1	1	1
Not specified	0	0	0	1	3	3	6	13	16	34
Medicine	2	4	6	11	13	65	79	122	194	276
Linguistics	0	0	0	0	0	0	0	0	2	2
Law	0	0	0	0	0	0	0	1	1	1
Information Sciences	0	0	0	0	1	5	5	5	8	9
Industry	0	0	0	4	4	5	6	8	8	14
Humanities, Arts and Social Sciences	0	0	0	0	0	2	2	2	3	3
Economics	0	0	0	0	0	5	5	5	8	11
Computer Science	0	0	0	1	1	12	13	20	26	29
Communication	0	0	0	3	3	8	8	8	15	22
Business	0	0	0	0	0	5	8	10	10	16

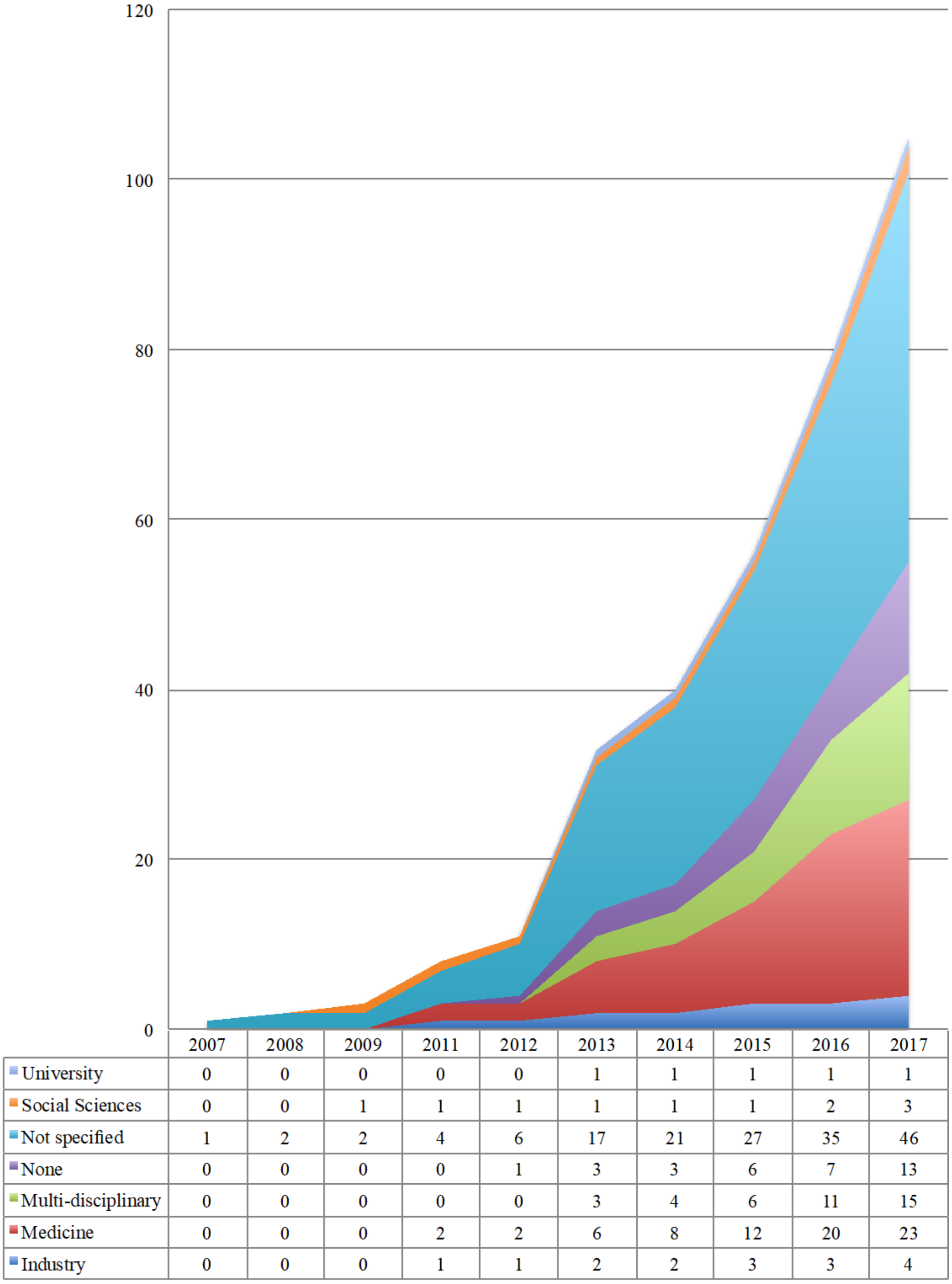
Medicine is a broad and diverse subject area and therefore a more granular analysis was conducted (Figure 13). This revealed that most of the studies (24%, n=66) affiliated to the medical subject area did not specify the sub-discipline from which they originate; however, the biomedical informatics (8%, n=22), psychology (8%, n=22), nursing (8%, n=23), surgery (6%, n=18) and internal medicine (6%, n=18) communities appear to be the most active to date.

Figure 13. Type of medical discipline to which the study authors were affiliated



Authors of the eligible NCD studies did not openly specify the sources of funding to conduct research analysing data from social media (44%, n=46), as can be seen in Figure 14. Those that did either did not receive any specific funding (12%, n=13) or received it from multi-disciplinary sources (14%, n=15), such as the National Science Foundation in the USA (nine studies) or the Economic and Social Research Council in the UK (three studies) or institutions that fund medical research, such as the National Institutes of Health (nine studies).

Figure 14. Cumulative volume, between 2007-2017, of eligible NCD studies by subject area of study funding



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
University	0	0	0	0	0	1	1	1	1	1
Social Sciences	0	0	1	1	1	1	1	1	2	3
Not specified	1	2	2	4	6	17	21	27	35	46
None	0	0	0	0	1	3	3	6	7	13
Multi-disciplinary	0	0	0	0	0	3	4	6	11	15
Medicine	0	0	0	2	2	6	8	12	20	23
Industry	0	0	0	1	1	2	2	3	3	4

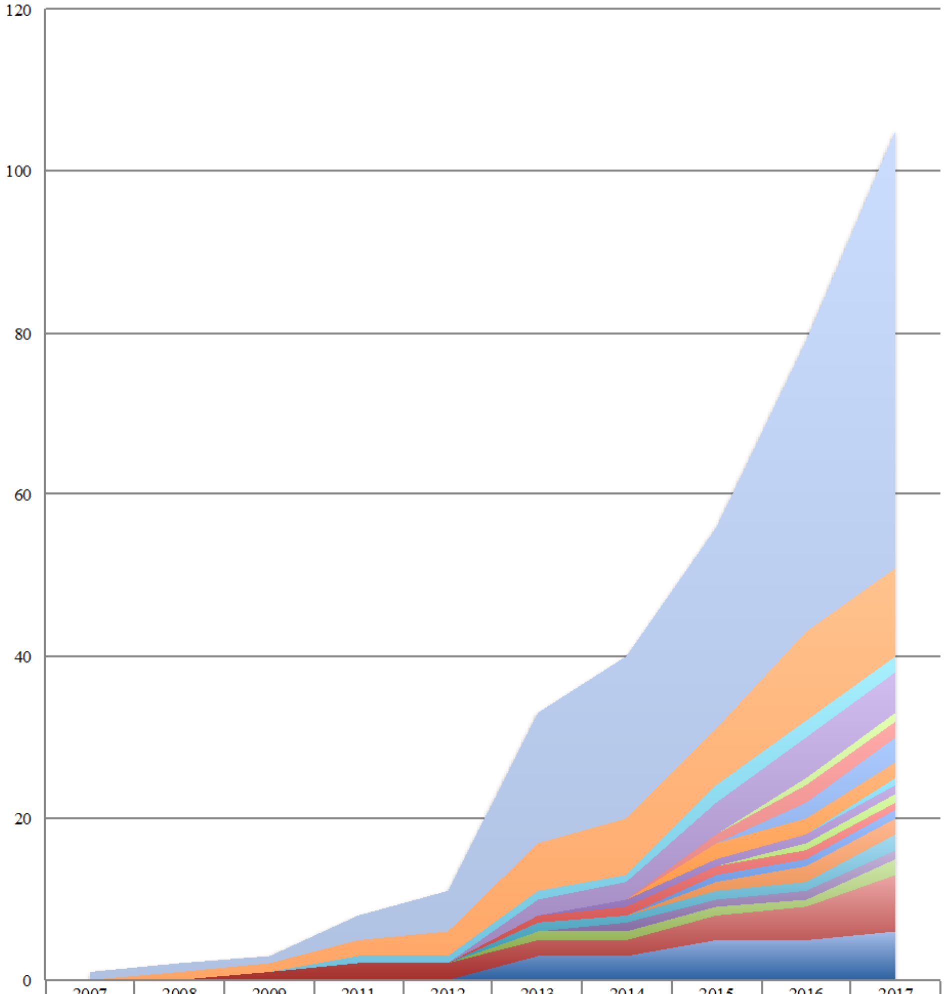
The journals and conference proceedings in which these studies are published vary by Scimago Journal subject area [96]. The subject areas were analysed by inputting the data into a text visualisation tool [205]. The resulting word cloud (Figure 15) was created from the 95 different words and maps the frequency of different terms appearing in the corpus of text. Words with a frequency higher than three are displayed. This highlights that the most popular journals and conference proceedings in which to publish research that analyses data from social media are those associated with medicine, health, computer science, informatics, nursing and miscellaneous topics.

Figure 15. Word cloud of journal subject areas for the 105 eligible NCD studies with frequency greater than three

Geographical scope of studies

The location of the first authors' affiliation was analysed to determine in which countries the eligible NCD studies originated. The cumulative volume, over time, by geographical location of first author affiliation is described in Figure 16 and 51% (n=54) of the current literature using data from social media originated in the USA, followed by UK (11%, n=12), Canada (7%, n=7), Australia (6%, n=6) and Sweden (5%, n=5).

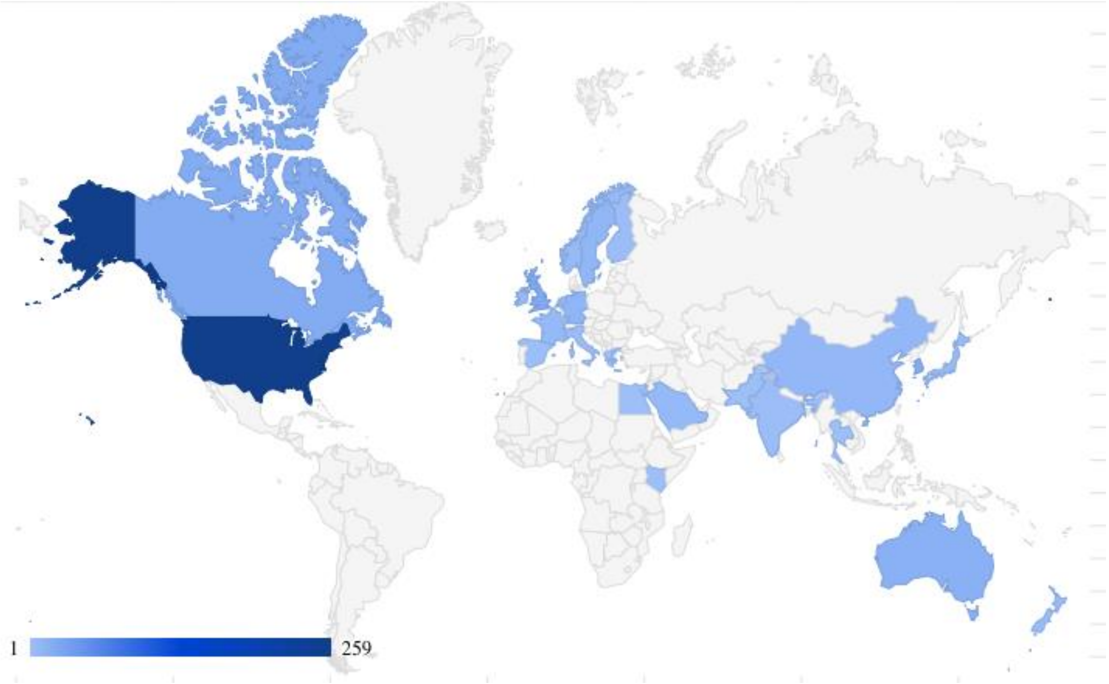
Figure 16. Cumulative volume, between 2007-2017, of NCD eligible studies by location of the first author’s affiliation



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
USA	1	1	1	3	5	16	20	25	36	54
United Kingdom	0	1	1	2	3	6	7	7	11	11
The Netherlands	0	0	0	1	1	1	1	2	2	2
Sweden	0	0	0	0	0	2	2	4	5	5
Spain	0	0	0	0	0	0	0	0	1	1
Saudi Arabia	0	0	0	0	0	0	0	1	2	2
Rep. South Korea	0	0	0	0	0	0	0	0	2	3
Norway	0	0	0	0	0	0	0	2	2	2
New Zealand	0	0	0	0	0	0	0	0	0	1
Japan	0	0	0	0	0	0	1	1	1	1
Italy	0	0	0	0	0	0	0	0	1	1
Ireland	0	0	0	0	0	1	1	1	1	1
India	0	0	0	0	0	0	0	1	1	1
Germany	0	0	0	0	0	0	0	1	2	2
France	0	0	0	0	0	1	1	1	1	2
Finland	0	0	0	0	0	0	1	1	1	1
China	0	0	0	0	0	1	1	1	1	2
Canada	0	0	1	2	2	2	2	3	4	7
Australia	0	0	0	0	0	3	3	5	5	6

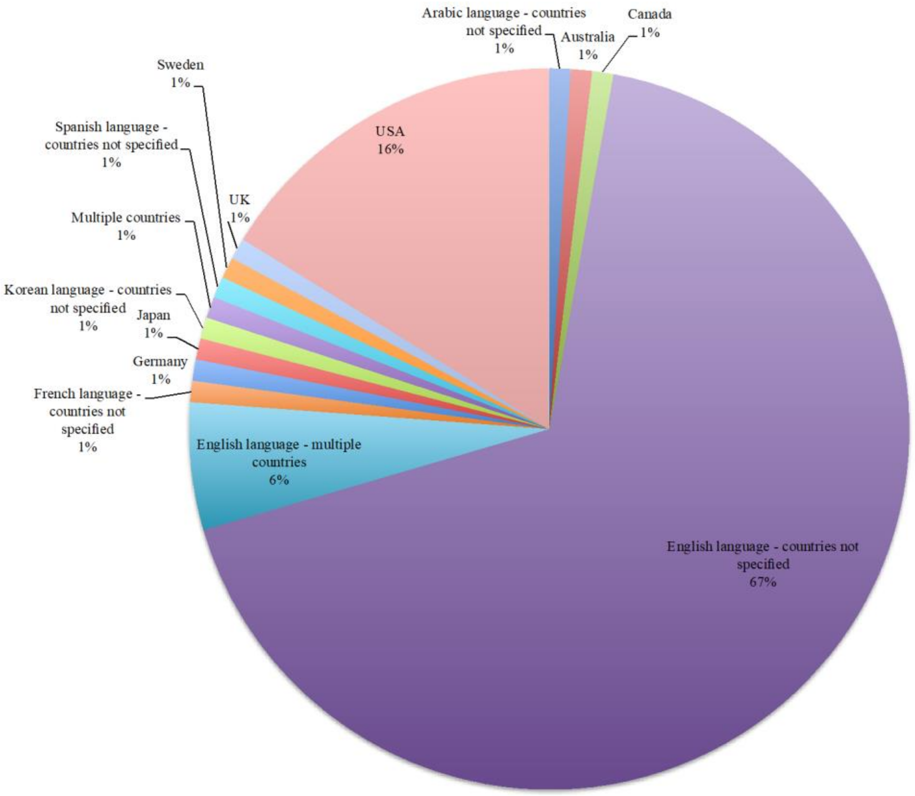
The location of the first author for each of the eligible NCD studies is reflected in Figure 16 however it is acknowledged that the relative contribution and etiquette associated with the order that authors are listed in a scientific paper can vary by discipline. It is worth noting that the average number of authors per study was 4.18 and therefore further analysis was carried out to include all authors of the 105 eligible studies. Ninety-one per cent (n=361) of the 397 distinct authors contributed to only one of the eligible studies, whilst 7% (n=26) contributed to two studies, 2% (n=6) to three studies and 1% (n=4) to four studies. The geographical locations of the affiliates of all named authors involved in the eligible NCD studies are represented in Figure 17.

Figure 17. Frequency of authors of the eligible NCD studies by geographical location of affiliate



The geographical scope of the sample of social media data analysed in the eligible NCD studies was also investigated. The results of which are illustrated in Figure 18. Due to the nature of the Internet and its general accessibility independent of traditional national borders [206], most of the eligible studies did not determine their sample collection methods based on geographical location, but instead used language as a key inclusion criteria. Ninety-two per cent of the studies sourced either English language posts or posts from countries where the predominant language was English. This includes 67% (n=71) of the studies using English language posts that did not specify a country, 16% (n=17) from the USA and 6% (n=6) from multiple English language-speaking countries. Eligible studies that analysed social media data in languages, such as Japanese (1%, n=1), Swedish (1%, n=1), German (1%, n=1), Spanish (1%, n=1), French (1%, n=1), Korean (1%, n=1) and Arabic (1%, n=1), were in the minority.

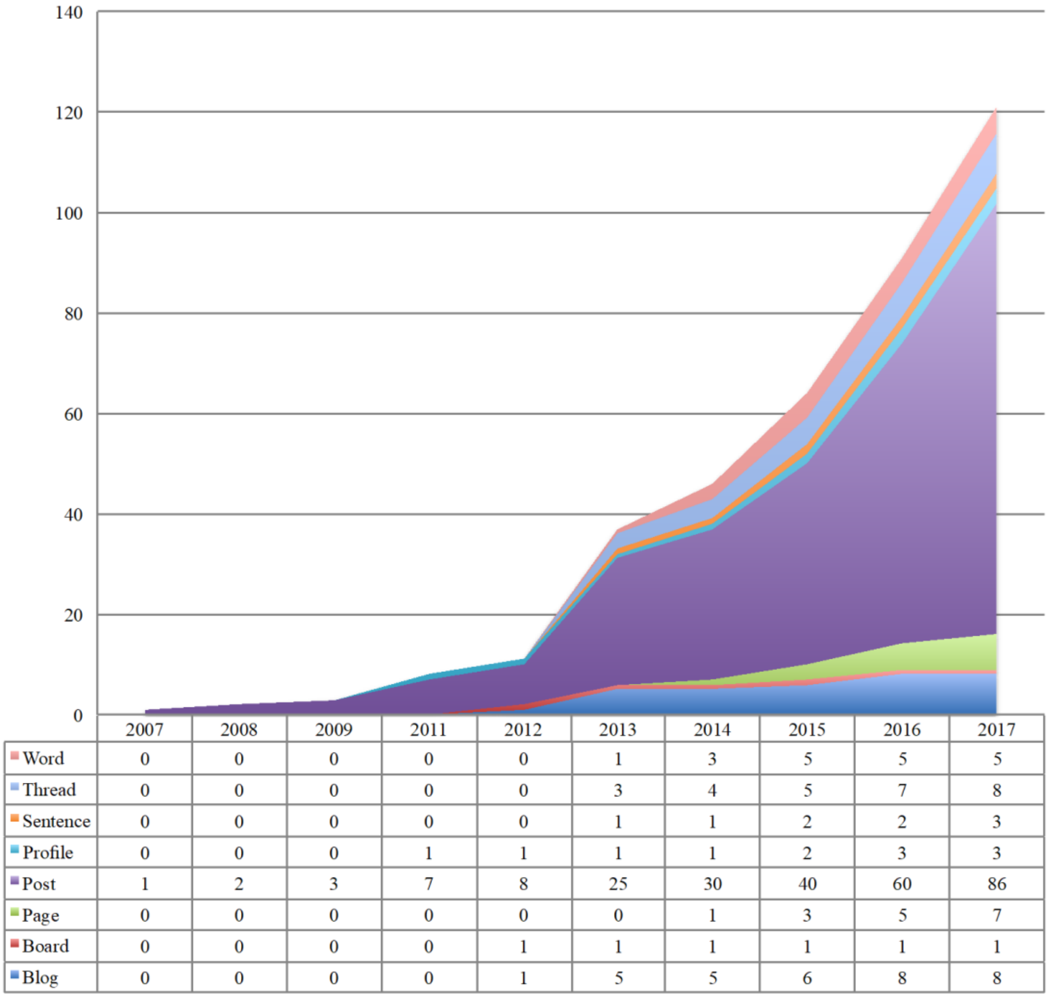
Figure 18. Percentage of eligible NCD studies by geographical location and language of social media data sample



Extraction and analysis of social media data

The units of analysis for the 105 eligible NCD studies were identified as either posts, boards, blogs, profiles, pages, threads, sentences or words with the number of studies analysing posts increasing between 2007-2017 (Figure 19). Fifteen of the studies analysed multiple unit types. The dominant unit of analysis was posts (71%, n=86), which included Tweets.

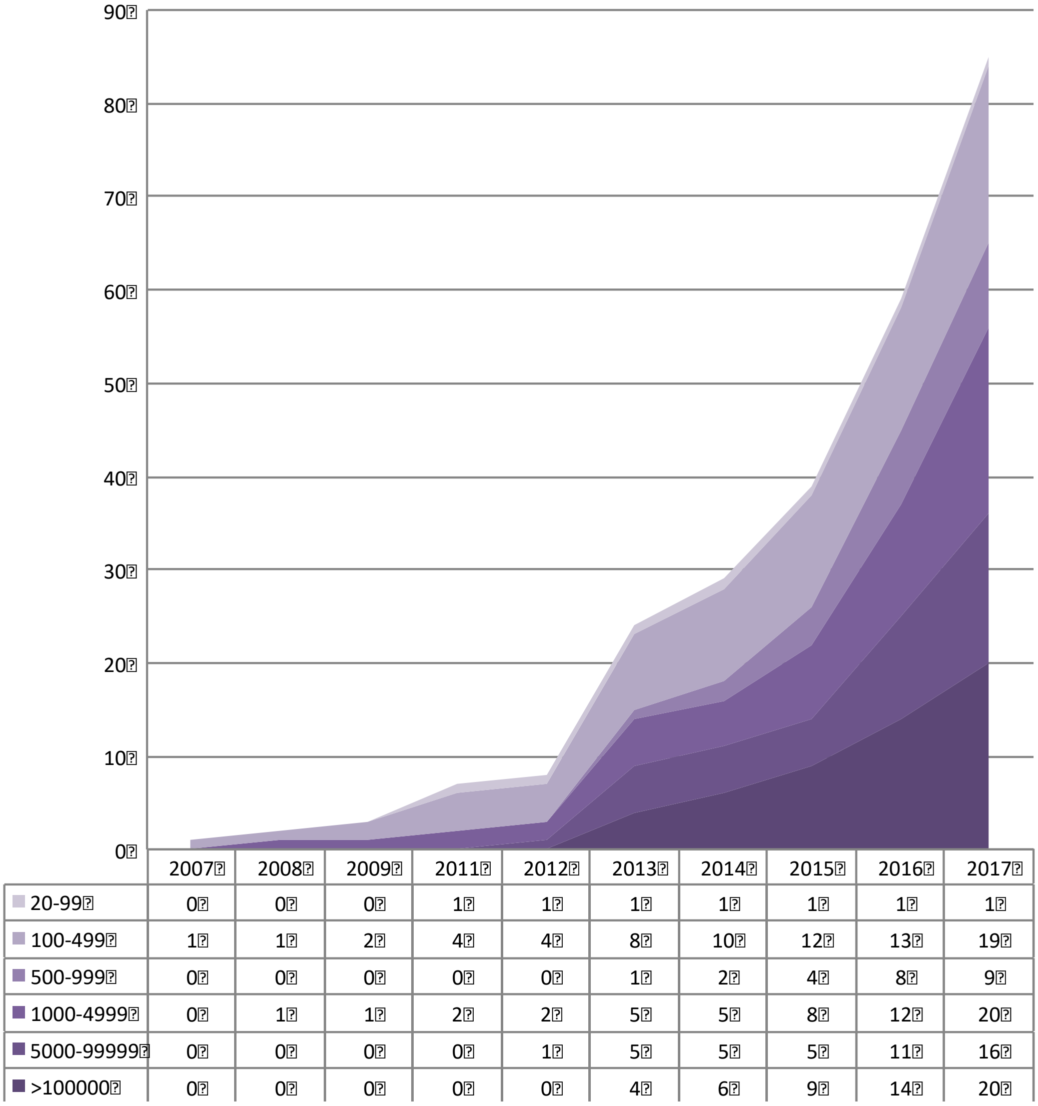
Figure 19. Cumulative volume, between 2007-2017, of eligible NCD studies by unit of analysis



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
Word	0	0	0	0	0	1	3	5	5	5
Thread	0	0	0	0	0	3	4	5	7	8
Sentence	0	0	0	0	0	1	1	2	2	3
Profile	0	0	0	1	1	1	1	2	3	3
Post	1	2	3	7	8	25	30	40	60	86
Page	0	0	0	0	0	0	1	3	5	7
Board	0	0	0	0	1	1	1	1	1	1
Blog	0	0	0	0	1	5	5	6	8	8

When analysing the 86 studies that extract and analyse posts, it was clear that the size of the sample analysed was increasing over time (Figure 20). Studies extracting and analysing samples with greater than 5,000 posts were published from 2012 onwards and account for 42% (n=36) of the 86 studies. By 2017, 23% (n=20) of eligible NCD studies had sample sizes of greater than 100,000 units.

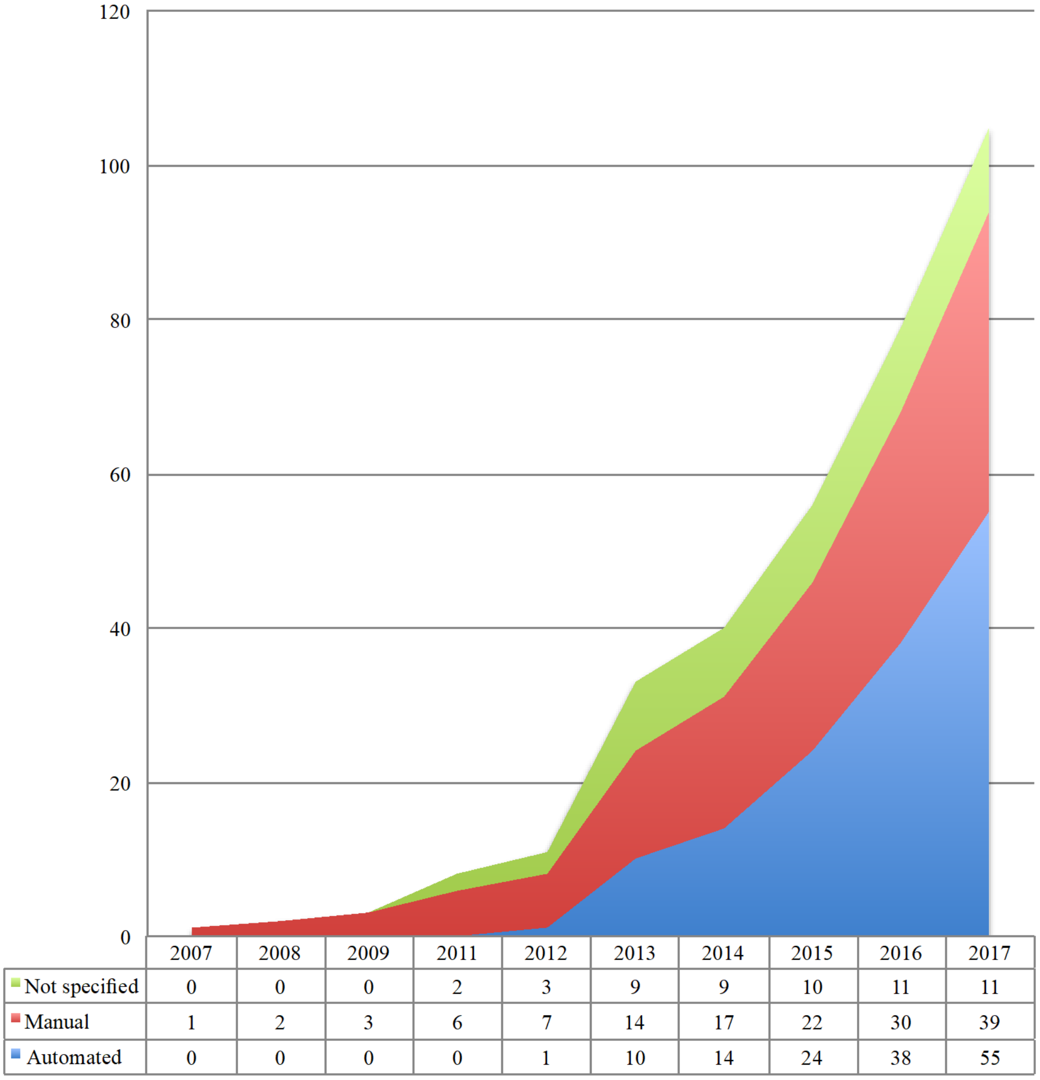
Figure 20. Cumulative volume, between 2007-2017, of eligible NCD studies by sample size of posts



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
20-99	0	0	0	1	1	1	1	1	1	1
100-499	1	1	2	4	4	8	10	12	13	19
500-999	0	0	0	0	0	1	2	4	8	9
1000-4999	0	1	1	2	2	5	5	8	12	20
5000-99999	0	0	0	0	1	5	5	5	11	16
>100000	0	0	0	0	0	4	6	9	14	20

The methods for extracting the data from the social media platform were categorised as either manual or automated. Figure 21 shows that earlier studies used manual methods and that the number of studies that applied automated methods had rapidly increased since 2012, in line with the increasing sample size of data analysed in the eligible NCD studies. The use of manual methods also increased; however, the trend suggests that this was slowing down.

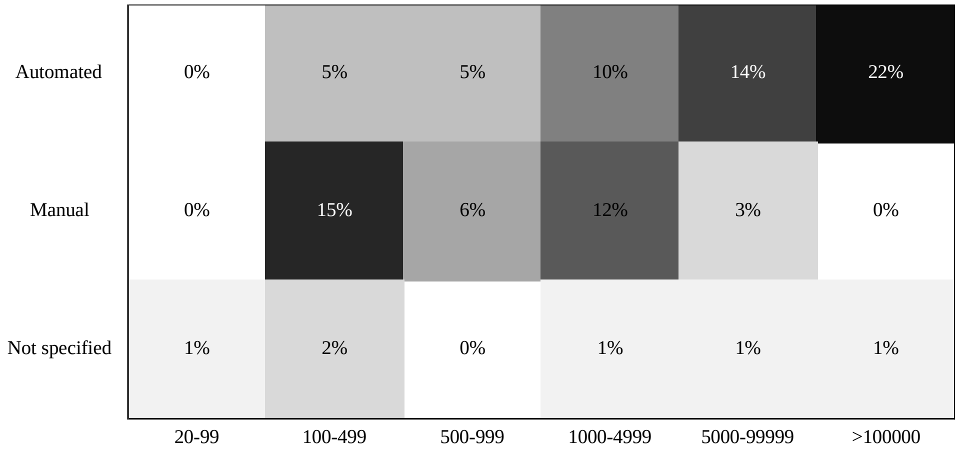
Figure 21. Cumulative volume, between 2007-2017, of eligible NCD studies by automated or manual method of extracting the data from social media



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
🟩 Not specified	0	0	0	2	3	9	9	10	11	11
🟥 Manual	1	2	3	6	7	14	17	22	30	39
🟦 Automated	0	0	0	0	1	10	14	24	38	55

A heat map of the number of posts versus the method of extraction (Figure 22) reveals that, once the sample size is greater than 5,000 units, automated methods of extraction are predominantly applied.

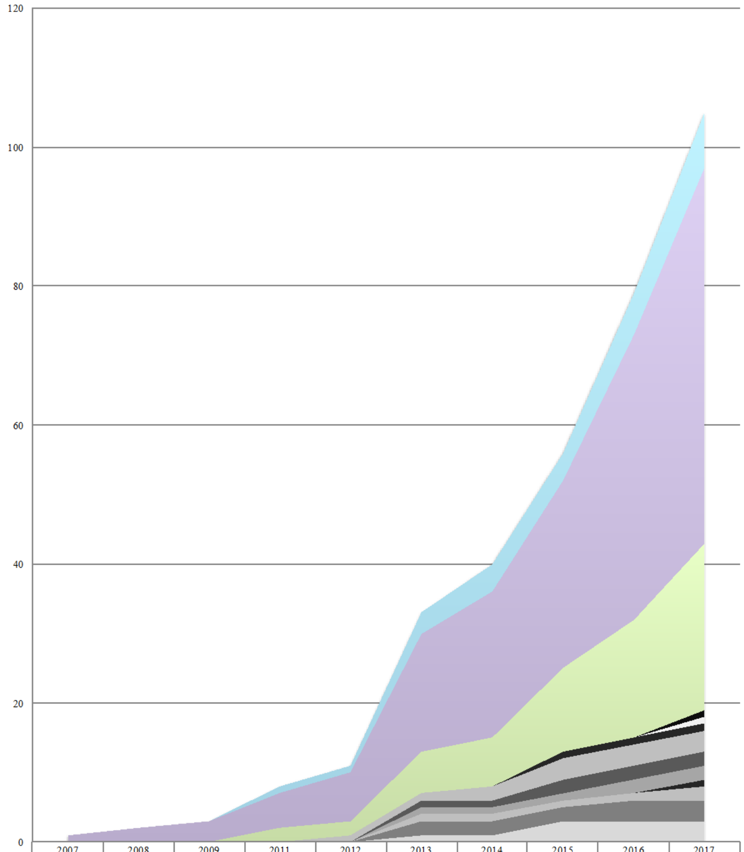
Figure 22. Percentage of eligible NCD studies, that apply automated and manual extraction methods in relation to sample size



	20-99	100-499	500-999	1000-4999	5000-99999	>100000
Automated	0%	5%	5%	10%	14%	22%
Manual	0%	15%	6%	12%	3%	0%
Not specified	1%	2%	0%	1%	1%	1%

Figure 23 shows how the use of different methods of extraction have evolved between 2007 and 2017, with the majority of the 105 eligible NCD studies using qualitative methods (51%, n=54), not specifying the extraction method (22%, n=24) or mixed methods (8%, n=8). The data mining techniques identified by Injadat et al. started to appear in 2012 (grey scale in Figure 23). Only 27% (n=15) of the 55 eligible NCD studies that used automated methods applied at least one of the data mining techniques identified by Injadat et al. (grey scale in Figure 23) and these focused primarily on Support Vector Machine (11%, n=6), Artificial Neural Networks (7%, n=4) and Bayesian Networks (7%, n=4). Twelve of the 19 data mining techniques did not appear at all in any of the eligible NCD studies. Latent Dirichlet Allocation (4%, n=1) and Linguistic Inquiry and Word Count (4%, n=1) were identified as two additional data mining techniques applied in the 55 eligible NCD studies that used automated methods for extracting the data from social media.

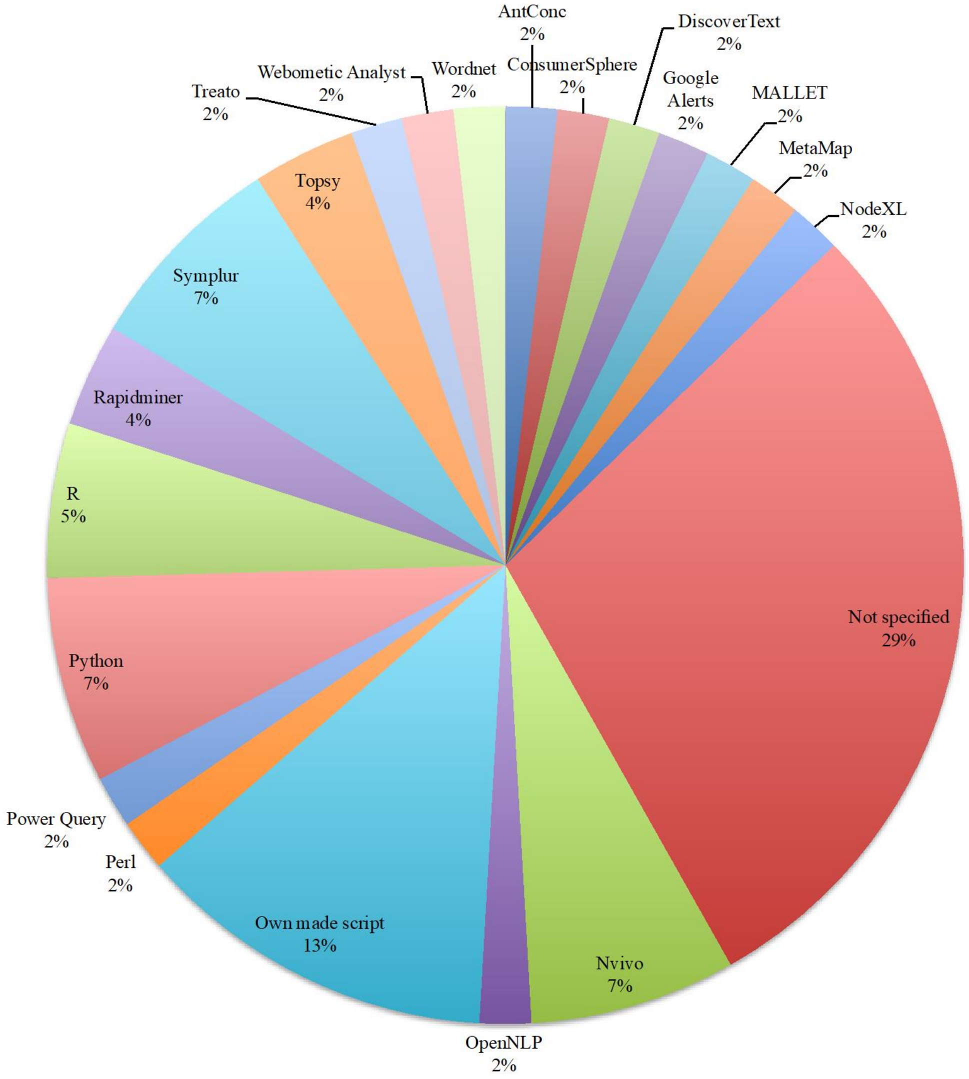
Figure 23. Cumulative volume, between 2007-2017, of eligible studies by method of data extraction from social media



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
■ Mixed method	0	0	0	1	1	3	4	4	6	8
■ Qualitative methods	1	2	3	5	7	17	21	27	41	54
■ Not specified	0	0	0	2	2	6	7	12	17	24
■ Support Vector Machine, Latent Dirichlet Allocation, Artificial Neural Network	0	0	0	0	0	0	0	0	0	1
Support Vector Machine, Decision Trees, Logistic Regression	0	0	0	0	0	0	0	0	0	1
■ Support Vector Machine, Bayesian Networks, Decision Trees	0	0	0	0	0	0	0	1	1	1
■ Support Vector Machine	0	0	0	0	1	1	2	3	3	3
■ Novel	0	0	0	0	0	1	1	2	2	2
■ Linguistic Inquiry and Word Count (LIWC)	0	0	0	0	0	1	1	1	2	2
■ Linear Discriminant Analysis	0	0	0	0	0	0	0	0	0	1
■ Latent Dirichlet Allocation	0	0	0	0	0	1	1	1	1	2
■ Bayesian Networks	0	0	0	0	0	2	2	2	3	3
■ Artificial Neural Network	0	0	0	0	0	1	1	3	3	3

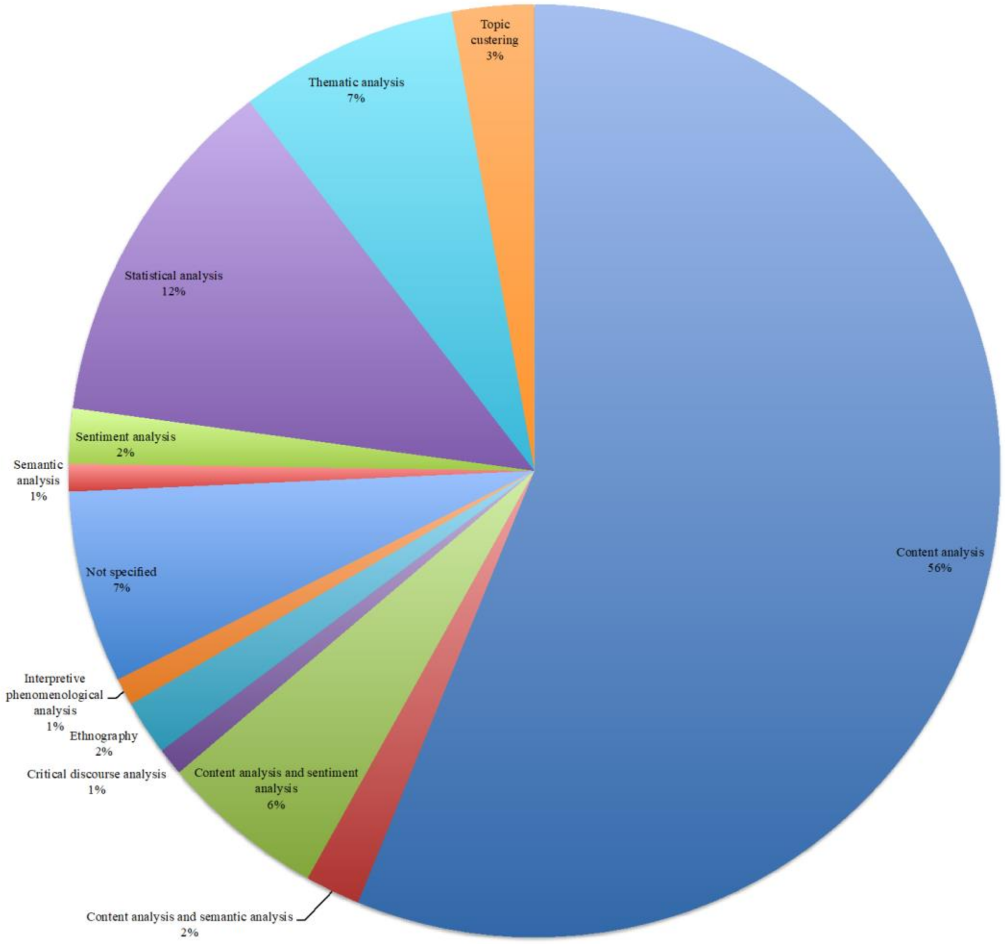
The tools, referenced in the eligible NCD studies, for the automated extraction of data from social media are represented in Figure 24. Most (29%, n=16) of the 55 studies did not specify which tool was used to automate the extraction of data; however, those that did had created their own script (13%, n=7) or used Symplur (7%, n=4), Python (7%, n=4), R (5%, n=3) or NVivo (7%, n=4). The first of these tools was referenced in 2012.

Figure 24. Percentage of propriety tools used, in the eligible NCD studies, to automate extraction of data from social media



In the 105 eligible NCD studies, the type of analysis performed on the data extracted from social media can be grouped based on those identified by Injadat et al. [98] and are represented in Figure 25: content analysis (64%, n=67), sentiment analysis (8%, n=8) and semantic analysis (3%, n =3). Additional types of analysis were evident in the studies and include statistical analysis (12%, n=13), thematic analysis (7%, n=8), topic clustering (3%, n=3), ethnography (2%, n=2), critical discourse analysis (1%, n=1) and interpretive phenomenological analysis (1%, n=1). Seven per cent (n=7) of the eligible NCD studies did not specify the type of analysis that was conducted.

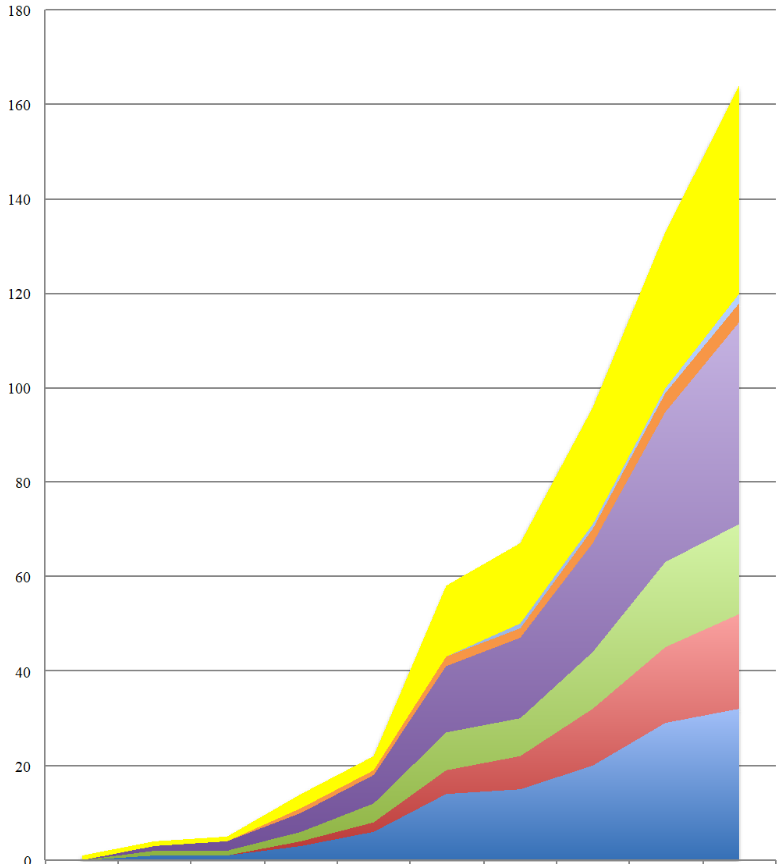
Figure 25. Percentage of types of analysis conducted in the 105 eligible NCD studies



Ethical considerations applied

An analysis of how ethical considerations are applied in the published literature from 2007-2017 (Figure 26) reveals that a large proportion (42%, n=44) of the eligible NCD studies did not make any reference to the 10 ethical concepts. Those that did mainly focused on IRB/regulatory approval (41%, n=43), privacy (31%, n=32) and/or informed consent (20%, n=20). None of the eligible studies referred to the ethical sensitivity of geographical information, the economic value of personal information, medical exceptionalism or the benefits of identifying socially harmful conditions.

Figure 26. Cumulative volume, between 2007-2017, of eligible NCD studies and their application ethical considerations (not mutually exclusive)



	2007	2008	2009	2011	2012	2013	2014	2015	2016	2017
🟨 No reference to any ethical considerations	1	1	1	3	3	15	17	25	33	44
🟪 Benefit of identifying socially harmful medical conditions	0	0	0	0	0	0	0	0	0	0
🟩 Medical exceptionalism	0	0	0	0	0	0	0	0	0	0
🟥 Economic value of personal information	0	0	0	0	0	0	0	0	0	0
🟦 Researcher lurking	0	0	0	0	0	0	1	1	1	2
🟧 Geographical information	0	0	0	0	0	0	0	0	0	0
🟧 Traditional research vs social media research	0	0	0	1	1	2	2	3	4	4
🟪 IRB / regulation	0	1	2	4	6	14	17	23	32	43
🟩 Ethical theory	0	1	1	2	4	8	8	12	18	19
🟥 Informed consent	0	0	0	1	2	5	7	12	16	20
🟦 Privacy	0	1	1	3	6	14	15	20	29	32

2.2.3 Discussion

The results of the systematic scoping review of literature reveal interesting insights into how this research field has evolved since the first social media platform was launched in 1997. Although the first identified study that extracted and analysed data from a social media platform to derive insights into health topics was published in 2002, it was not until 2007 that a study investigating NCDs was published. Since then the volume of literature has grown significantly. Despite this, only 15% (n=105) of the 695 eligible studies identified focus on NCDs. Further analysis on the types of NCDs investigated, the communities contributing to this growing body of knowledge, the geographical scope, as well as the sources of data, the ethical considerations applied and the methods for extracting and analysing the data highlights how the research field is changing and how it can be used to inform future studies.

The type of NCDs investigated in the existing literature

Cancer was identified as the most commonly researched NCD, followed by diabetes, a finding that is consistent with the results of Patel et al's 2015 systematic review [86] and aligns somewhat with the distribution of NCD research funding by the NIH [207]. Cancer Research UK, however, has identified over 200 different types of cancer each with its own causes, symptoms, treatments and care needs and that patient illness trajectories can vary depending on the severity and type of cancer experienced [208]. At the time of writing, there was no known systematic review that had sought to understand how the narratives associated with the different types of cancer are represented in social media. The apparent lack of differentiation between cancer types goes some way to explain why 40% (n=23) of the studies identified did not specify which cancer was under investigation and hence limits the conclusions that can be drawn from these studies. Those that did differentiate focused mainly on breast cancer [101, 123, 124, 131, 134, 136, 174-176, 187, 191, 198, 202, 204] and colorectal cancer [108, 109, 122, 185], aligning somewhat, with the exception of lung cancer, to the list of most common cancers provided by the WHO [209]. In the studies that examine multiple cancers (14%, n=8), five focus on two cancer types [111, 115, 116, 169, 173] and are gender specific; breast, cervical, prostate or testicular cancer, aimed at investigating preventative public awareness and screening campaigns such as Movember, pap smears and mammograms. Two of the studies compare four cancer types [113, 126] whilst one study examined 10 different types of cancer, finding that breast cancer, leukemia, uterine and cervical cancer, lung cancer, colon cancer and stomach cancer were the most commonly referenced cancers on Twitter in Japan [199]. Despite their prevalence in society, cancers, and the funding available to conduct research that investigate them, have not yet attracted the attention of the research community analysing data shared on social media. Examples of this include lung cancer (5%, n=3) and leukemia (2%, n=1). This, combined with the discrepancy between the types of cancer studied in the published literature and those represented in the social media narratives, highlights gaps in the current landscape that warrant further research.

Similar conclusions can be drawn from the 20 studies [100, 102, 103, 105, 106, 112, 121, 138, 143, 147-153, 163, 167, 180, 189, 203] that investigate diabetes content on social media, as only three [138, 163, 180] of these make a distinction between the types of diabetes (Type 1, Type 2 and gestational). Other NCDs that are under-represented in the existing literature are those associated with chronic respiratory

(2%, n=2) and cardiovascular conditions (7%, n=7), hence creating further avenues for future empirical research.

Academic communities most active in this field of research

Gruzd published an automated content analysis of 14,500 abstracts of existing literature on social media, in 2015 [210]. Preliminary findings from that research identified health as being one of the main topic areas. NCD-related sub-topics, which were identified as having higher frequency, included cardiovascular, colorectal cancer and diabetes management. Although there are some overlaps with the findings from this systematic scoping review, there is a lack of transparency into Gruzd's research design and results, suggesting that conclusions from his study are limited.

Insights into the academic communities that are most active in extracting and analysing NCD-related content from social media can be derived from the author's affiliations, the institutions funding this type of research as well as the journals in which the studies are published. Four hundred and thirty-five unique authors conducted the 105 eligible NCD studies and are affiliated to 259 institutions. Sixty-two per cent (n=161) of these institutions are self-reported as medical schools, whilst only 7% (n=18) are self-reported as computer science or technology departments (6%, n=14). Despite the fact of 8% (n=22) of the authors from the medical affiliates self-reporting their affiliation to biomedical informatics departments, this significant gap highlights the scale and prominence of siloed research despite its multi-disciplinary nature – a finding which is compounded by 14% (n=15) of the eligible NCD studies being sponsored by multi-disciplinary funding bodies misaligning, with the widely proclaimed value and benefit to society [211, 212] and patients [213-215] of conducting multi-disciplinary research that addresses complex, modern challenges [216] and the influence of university, industry and government (i.e. ' The Triple Helix') on propagating innovation [217]. It is recognised to be a difficult mode of operating [218] given the intellectual and social structures through which knowledge has traditionally been structured [219].

The 105 eligible NCD studies were published in 84 unique journals and conference proceedings, with the predominant Scimago Journal subject area being Medicine. Eighty-eight per cent (n=74) of these journals and conference proceedings contained only one eligible study. Eleven of the eligible NCD studies [109, 137, 146, 149, 153, 162, 166, 170, 172, 199, 201] were published in the Journal of Medical Internet Research (Scimago Journal subject areas: Medicine and Health Informatics), which claims to be 'the leading peer-reviewed journal for digital medicine and health and health care in the internet age'

[220]. Launched in 1999, its editor-in-chief is Gunther Eysenbach, who introduced the terms ‘infodemiology’ and ‘infoveillance’ in 2004 [27].

Nine of the eligible NCD studies [102-105, 110, 117, 121, 167, 179] were published in proceedings from IEEE conferences. The IEEE claims to be the ‘world’s largest technical professional organization dedicated to advancing technology for the benefit of humanity’ and the ‘trusted voice for engineering, computing and technology information’ [221]. Other journals, each with four eligible NCD studies, include the *Journal of Cancer Education and Cancer Nursing*, reinforcing the propensity for cancer-related studies.

Ethical considerations applied

In Gustafson and Woodworth’s 2014 review of the ethical issues in research using social media, the six studies identified did not sufficiently consider the risk of privacy incursion and did not seek ethical approval from an IRB to conduct the research [74]. Despite most of the eligible NCD studies having been conducted by researchers affiliated to medical institutions and being published in medical journals, it is surprising that 42% (n=44) of these did not take into account the ethical rigour that is expected of this discipline. This is noteworthy as, despite Version 1.0 of the ethical guidelines having been published by the Association of Internet Researchers (AoIR) in 2002, then subsequently updated in 2012 and 2019 [222] hence demonstrating the expansion and evolution of the topic, they 5
were not referenced in any of the identified NCD studies and very few studies went further than acknowledging consultation with their IRB, which is primarily undertaken for instrumental reasons. The dominance of instrumental over ethical considerations seen in the scientific papers reviewed suggests that researchers extracting and analysing data from social media are heavily dependent on IRBs and journal editors to play the role of their ethical conscience. Researchers submitting articles to the *Journal of Medical Internet Research* are provided with a detailed list of ethical requirements ranging from privacy to IRB approval, the study of human subjects and the need for informed consent [220]. However this, unfortunately, is not the case for all journals and conferences, such is the case for the IEEE guidelines, which focuses on publication ethics [223], including citations, data accuracy and plagiarism, rather than ethical considerations for study conduct. It is therefore essential that ethics committees and editors evaluating research using social media data are aware of the range of platforms available and how they work, and can draw on the latest multi-disciplinary guidelines to inform their decision-making. In order to address this gap, further research is warranted into the ethical principles and guidelines available

to researchers who extract and analyse data from social media, so as to avoid ethically questionable or even illegal practices, such as was the case with Facebook-Cambridge Analytica in 2018 [224].

Sources and methods for extracting and analysing data from social media

The different sources of data, unit of analysis, sample size, methods of extraction and analysis, as well as the propriety tools applied, provide further insight into the research design of the 105 eligible NCD studies. Micro-blogging sites (e.g. Twitter), social networking sites (e.g. Facebook), discussion forums and blogs are the main sources of data, aligning somewhat with the findings from Patel et al.'s 2015 systematic review, which identified Facebook, blogs, Twitter and wikis as the most commonly studied social media [86]. Research from the Pew Research Center identifies YouTube and Facebook as the most-widely used online platforms, with a smaller share using Twitter, Pinterest, Instagram and LinkedIn [225]. Despite inconsistencies about which social media are most commonly used for research purposes, it is evident that mainstream social media rather than condition- or health-specific platforms are the dominant source of data in the published research and can be attributed to their reach, the volume of active users worldwide, the volume of public real-time content and the ease of access and searching [83].

Access to large data sets can be cost-prohibitive and time-consuming for many researchers. To overcome this, Twitter goes so far as to promote access to data for research purposes and innovation [226]. The use of hashtags for targeted Tweet topic identification and searching led Symplur, a social media analytics company, to initiate the Healthcare Hashtag Project [227]. According to Sinnenberg et al. [83], data sourced from Twitter for research purposes is used predominantly for content analysis, surveillance, engagement and social network analysis, with 41% of the studies using the Twitter application programing interface (API) to mine Twitter and others using NodeXL, Topsy and NCapture. Although identified in this systematic scoping review as propriety tools used in the eligible NCD studies, they were in the minority and were applied in only one [181], two [125, 159] and none of the studies, respectively.

Although Facebook has a Research division [228] to engage academic institutions and researchers and provides tools for developers [229], access to the data shared by users of this social media is restricted for commercial purposes and privacy reasons. Therefore, research extracting and analysing data from Facebook is limited to content in public profiles, pages or groups. It is interesting to note that, of the 16 eligible NCD studies that source data from Facebook [100, 101, 106, 125, 128, 135, 139, 143, 146, 158, 163, 188, 189, 195-197], seven are investigating cancer [101, 135, 139, 158, 195-197] and five diabetes

[100, 106, 143, 163, 189], aligning with the earlier-discussed dominance of these two NCDs in the current literature.

The use of discussion forums as a source of data in the eligible NCD studies is often associated with popular health-related websites, such as Breastcancer.org [134, 202, 204], DiabetesDaily.org [102, 105], Medhelp.org [171, 172, 176] and WebMD.com [111, 152, 153, 179]. Medhelp.org provides medical and health information about multiple conditions and has 155 online communities [12]. WebMD.com is an American online publishing corporation that provides health-related information through a website, magazines and applications [230]. Their websites include message boards and blogs for members to connect and share their experiences. Breastcancer.org, on the other hand, is a non-profit organisation that provides information, research news and online support through the world's largest online breast cancer community of more than 226,000 members [231]. DiabetesDaily.org provides similar information about diabetes and claims to be the leading online community, with almost 1 million members [232].

Seven of the eligible NCD studies extracted and analysed data from multiple social media [118, 125, 128, 141, 158, 177, 200]. Two studies sourced data from two different social media [141] [158], one from three [125], one from four [128], one from eight [118], one from 10 [200] and one from over 3000 social media [177]. None of these studies went so far as to explain the differences in digital architecture, utility and affordances offered by these social media and the value these platforms have for people living with different conditions. Further research is therefore warranted.

Figure 20 highlighted the increasing volume of eligible NCD studies that extract and analyse large data sets. As can be seen from the heatmap (Figure 22), in most cases studies of over 5,000 units used automated methods of extraction, whilst studies with less than 500 units were extracted manually. Manual methods for extraction of data for the eligible NCD studies were common until 2012, when automated methods e.g. Support Vector Machine, started to appear in the NCD literature [120].

Terms such as 'big data' and 'data mining' have been around since the 1990s to describe the Volume, Velocity, Variety, Veracity and Value of data, as well as automated methods for extracting and analysing it [98, 233-235]. Findings from this systematic scoping review highlight that Support Vector Machine, Artificial Neural Networks and Bayesian Networks are the most common data mining techniques represented in the literature, which is in line with previous research [98]. However, it is interesting to observe that 12 of the 19 techniques for mining data from social media, as described in Injadat et al's 2016 study, do not appear in the eligible NCD studies and that two additional techniques were identified, therefore highlighting the relative infancy and evolution of this field and the need for further research.

The benefits of extracting and analysing health data from social media were identified as the speed of generating more available real-time health information, access to hard-to-reach population segments and the potential for public health surveillance, to name a few [75]. Despite these potential benefits, methods of automation present their own challenges and limitations when deducing insights from data sourced from social media. These include: the fragmented landscape of contributors to the field; the ethical considerations [236] as well as the reliability of algorithms when applied in different contexts; the potential amplification of existing biases and therefore lower accuracy for minority groups who are underrepresented in the data set; the accuracy of text classification; inter-coder reliability challenges; and the shortfalls in interpreting meaning from text when compared to human ability [237]. Despite having sparse context links, the strength of the Support Vector Machine method is that it is one of the recognised techniques for solving classification problems [98] and could explain why it was applied in 11% (n=6) of the 55 eligible NCD studies that used automated methods for extraction of the data from social media. As the volume of studies continues to increase and the different research disciplines converge, it will be important to monitor and understand the evolution of the methods, tools and techniques used so as to provide recommendations for informing future research.

Geographical scope of studies

Results from the analysis of the geographical scope of studies reveal that the first author of 97% of the eligible NCD studies is located in high-income countries [97] such as the USA (51%, n=54), UK (11%, n=12), Canada (7%, n=7), Australia (6%, n=6) and Sweden (5%, n=5). The only higher-middle income countries and lower-middle income countries from where the first author is affiliated are China [109, 171, 172] and India [179], respectively. When analysing the location of all authors associated with the eligible NCD studies, additional countries were represented. These countries included Israel [128], Kenya, Pakistan, Egypt [150] and Thailand [164]. Scimago's ranking of countries by number of publications lists the USA ranked number 1, followed by China (ranked second), UK (ranked third), Canada (ranked seventh) and India (ranked ninth) [238]. This differs somewhat from countries investing in academic research and innovation, into which this field of research is justified, where the USA is ranked 28th and UK ranked 26th [239, 240]. A number of drivers and influences can explain this variation; however, further investigation is required to verify their impact. These drivers include: the number of research institutions per country [241] and the research disciplines on which they focus [242]; variations in Key Performance Indicators by which different institutions and research disciplines measure and value the performance and impact of individuals within the research community [243]; access to funding to conduct the research and the cost- prohibitive fees associated with publication [244]; and the prevalence of certain medical conditions within each country and their societal and economic impact [245].

The number of users of social media is also increasing [246], with distinct variations between countries [247]. As of 2016, some social media have a larger number of worldwide account holders than countries have population size [248]. China, India, USA and Brazil are identified as countries with the highest numbers of users of social media [249]; however, few of these countries are represented in the results of the eligible NCD studies. Results from this systematic scoping review revealed that English-language and English-speaking countries have greater representation in studies that source data from social media platforms (92%). This finding highlights the challenge and limitation of conducting this type of research, due to the existence of a ‘digital divide’ [250] that excludes those that do not have access to the Internet, whether it be due to geographical location, availability of skills or other barriers to digital inclusion [251]. Although research exists into the ‘digital divide’ and its consequences for research that source data from social media [252-254], little is known about its implications for studies that investigate NCDs.

Limitations

Despite going some way towards addressing the deficiencies of earlier systematic, quasi-systematic and scoping reviews and providing direction for future research, the preliminary results of the systematic scoping review of literature have their own limitations. These are mainly associated with the research design. The searches of the electronic databases were conducted in 2018 and, as per the published protocol the timeframe for inclusion of literature in this systematic scoping review was 1997-2017. Studies published at later dates (i.e. 2018-2020) were not included despite the increasing volume of literature that sourced data from social media. This is therefore considered a limitation of this systematic scoping review and the associated findings. The structured search of five electronic databases was intended to maximise the likelihood of identifying potentially relevant studies, whilst acknowledging that many of these would be excluded at later stages. Due to the extraordinary volume of identified peer-reviewed literature (13,008), the decision was made to diverge somewhat from the published protocol [255]. The omissions included: 1) 'snowballing' the article reference lists of the 605 eligible studies; 2) the literature identified using the Google search engine; 3) focusing the classification of the literature on the subset of NCD-related research rather than the broader portfolio of health topics identified in the WHO's 17 Sustainable Development Goals [256]; and 4) a second researcher independently cross-checking and ensuring the appropriate application of the inclusion and exclusion criteria and classification frameworks; as well as 5) only including studies that extracted and analysed text content, rather than content from rating sites, virtual worlds, images or videos from social media platforms such as YouTube, Flickr or Instagram, unless the text comments associated were analysed.

The electronic databases that were searched, as well as the publication language of the literature extracted, were limited to English and therefore may explain the high percentage of studies that sourced English language posts from social media. Further research would benefit from sourcing literature from Chinese, Spanish, French and Arabic databases. It is also worth considering the impact of multi-morbidities, as in society it is not uncommon for an individual to be diagnosed and living with multiple NCDs or a combination of an NCD and another condition [257]. .

Aside from further manual analysis to address these limitations and to complete the scoping review described in the published protocol by extending it to 2020, automated methods for conducting systematic reviews [258, 259] are being introduced that reduce the cost of production of such reviews, limit human error and enable the timeliness of insights despite the increasing volume of literature available. These approaches also have their challenges These include the variability of the databases to which they are suited, the need for human intuition and expertise to perform certain tasks within the process and the interpretations of the results [260].

2.2.4 Conclusions

The results of this comprehensive, systematic scoping review of NCD-related research using social media highlights the evolution of the field and provides insight into the narrative of different NCDs and their risk factors, types of methods used for extracting and analysing data from social media, the ethical considerations of conducting such research and the different social media used as sources of data.

Automated methods of extraction and analysis

Although automated methods for extracting and analysing data from social media were applied in other health-related studies, such as those focused on infectious disease surveillance, these methods only started to make an appearance in the literature investigating NCDs in 2012 and have been increasingly used. The variations in efficacy of these methods warrant further investigation to inform the research design of future studies.

Ethical considerations

Most of the identified NCD studies were produced and published by the medical community. Despite ethics guidelines, such as those published by the AoIR and the BPS [261, 262] being available, these were not referenced in any of the identified NCD studies. The lack of rigorous consideration given to the ten ethical concepts, beyond acknowledging consultation with IRBs, points to a need for further research into the ethics of conducting studies that source data from social media..

Types of NCDs investigated and the different sources of data

The results from this systematic scoping review highlight that cancer and diabetes dominate the published literature. In 2013, 18% (n=6) of the eligible studies focused on diabetes. This increased to 20% (n=21) of the 105 studies by 2017. Only three of these, however, investigated the differences between diabetes types [138, 163, 180] and therefore further research into the utility and affordances offered by different mainstream and condition-specific social media to those living with Type 1 diabetes is warranted.

Despite cancer being the topic of investigation in 54% (n=57) of the eligible NCD studies, lung cancer was only represented in three of the 105 eligible studies. These three studies sourced data from multiple discussion forums [104], Twitter [159] and an unspecified platform [168]; however, as was the case with diabetes Type 1, none of the studies drew comparisons between different social media. Further research is warranted to address this gap.

94% (n=98) of the 105 eligible NCD studies are cross-sectional in nature, analysing a sample of data sourced from social media at a specific point in time, from participants with varied characteristics. This has the benefits of containing multiple variables at the time of data capture, providing insights at a societal rather than individual level and creates opportunities to inform new in-depth research [263, 264] . The disadvantages of cross-sectional studies are that they cannot be used to analyse changes in behaviour over a period of time and they do not help determine cause and effect. Longitudinal studies, on the other hand, analyse data from the same sample over a period of time in order to understand changes at an individual level. Six of the studies identified in the eligible NCD literature were longitudinal in design [137, 157, 163, 167, 170, 204]. These studies investigated disease progression amongst adolescent cancer patients through analysing the content of blogs [157], monitoring outcome [137] and adverse events [167] experienced by patients taking different drugs, evolution of discussion topics in online cancer communities [170, 204] and the online activity of 20 diabetic patients through analysing the role of Facebook in the lives of these individuals [163]. None of the 105 eligible NCD studies sought to examine whether the illness trajectories used to inform planning of palliative and supportive care for patients who are terminally ill, were reflected in their digital narratives.

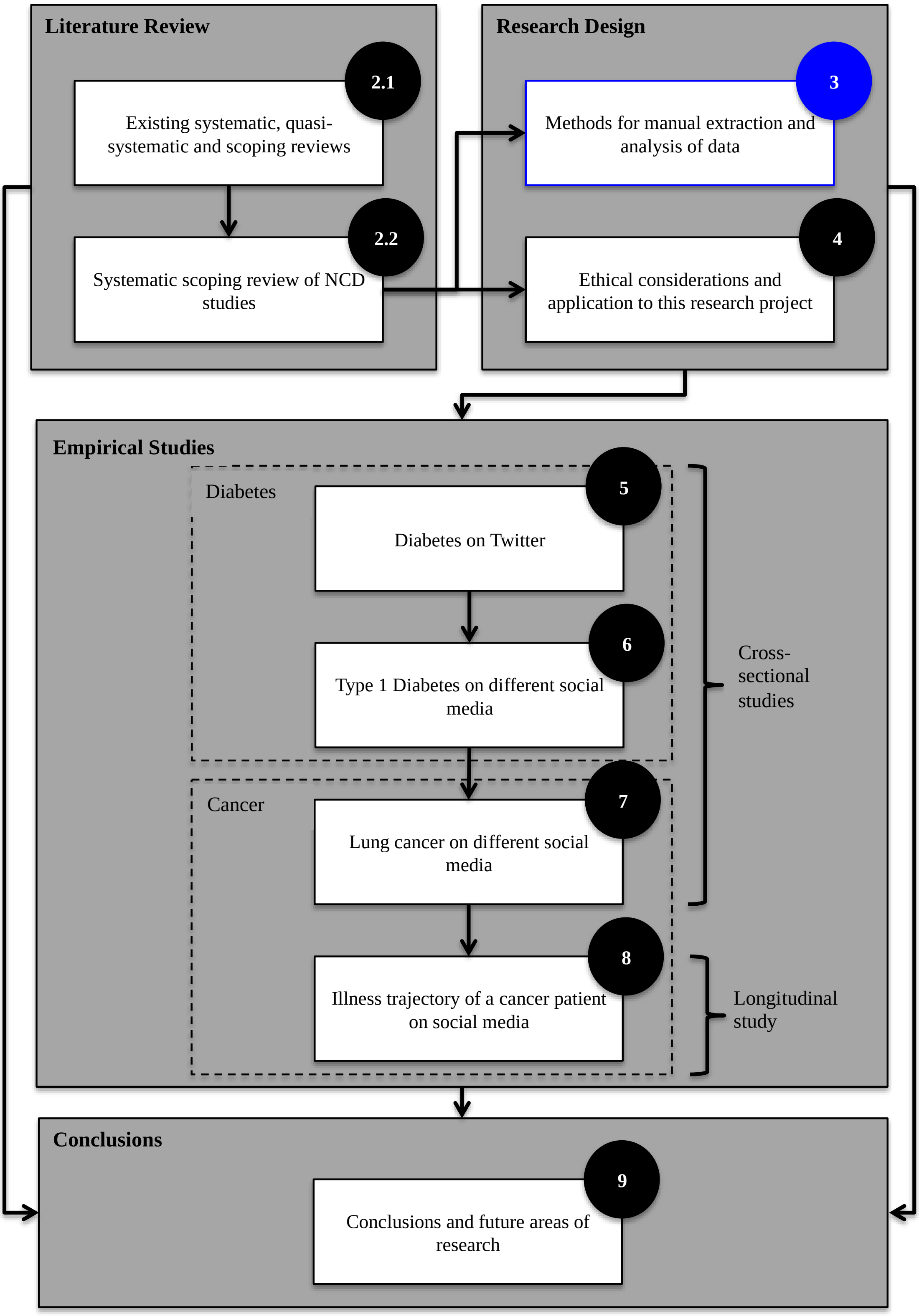
Despite the growing body of literature that source data from social media, this research remains siloed and fragmented. Although this systematic scoping review provided insights into the status of existing multi-disciplinary research literature that analyses data posted on social media for health research, further empirical studies are needed to better understand the narrative of different NCDs and their risk factors, the manual and automated methods for extracting and analysing data from social media, the ethical guidelines to be applied when conducting such research and the differences in affordances and utility offered by different social media depending on the medical condition.

2.2.5 Implications for this PhD research project

The studies conducted between 2013-2017, as part of this PhD research project and described in subsequent chapters, aimed to address the gaps in the existing literature. The manual methods and ethical considerations informed the research design of the empirical studies into the medical conditions of diabetes and cancer as well as the different social media from which data were sourced.

Chapter 3 discusses whether the PRISMA guidelines traditionally applied in systematic reviews can be re-purposed for analysing data from social media, whilst the ethical guidance available to researchers who extract and analyse data from social media are discussed in Chapter 4 and applied to the empirical studies described in Chapters 5-8.

Using data available from Twitter, which as previously discussed is easily accessible, the pilot study described in Chapter 5 provides preliminary insights into the differences between types of diabetes represented on social media. Type 1 diabetes was the sole focus of one study [138] in the corpus of published literature. The study was published in 2015 and sourced data from Twitter; as such, the conclusions that could be drawn were limited in their application to other social media. Further research into the utility and affordances offered by different mainstream and condition-specific social media to those living with Type 1 diabetes are described in Chapter 6 as well as for lung cancer in Chapter 7. The empirical study described in Chapter 8 provides early findings that address the deficit of research into whether the illness trajectories of terminally ill patients are reflected in their digital narrative.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

3. Can existing systematic review methods be re-purposed and applied to analyse data posted on social media?

As highlighted in the results of the systematic scoping review of the literature (Section 2.2), multiple manual and automated methods for extracting and analysing social media data are available and have been used to inform the research design of empirical studies. Despite the increasing number of public health research studies that extract and analyse this data, there is limited guidance on which manual methods to use in order to ensure repeatability and quality of results.

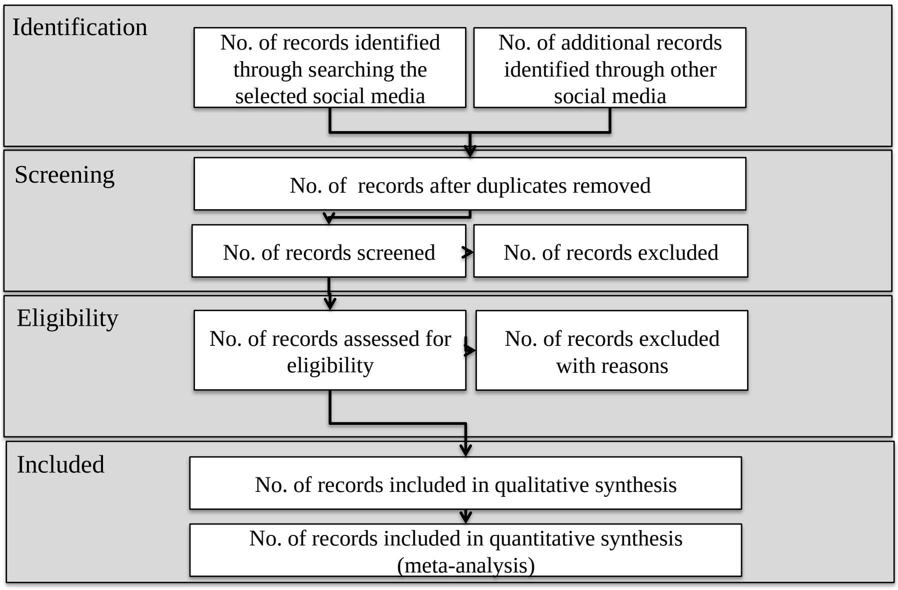
3.1 Background

Originally established in 1999 as QUOROM/PRISMA, this international group aimed to address the sub-optimal quality of reporting on randomised controlled trials. In 2009, the guideline was updated to address several conceptual and practical advances in the science of systematic reviews and was renamed as PRISMA (Preferred Reporting Items of Systematic reviews and Meta-Analyses) [265]. The group has since been acknowledged, within public health research circles, as an effective means of enabling a structured and robust review of literature, with records indicating a significant increase in the number of systematic reviews published, from 2,500 in 2004 to over 8,000 by 2014 [266]. The Cochrane Collaboration defines a systematic review as ‘a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review’ [267: pg45]. Although the definition refers to peer-reviewed scientific literature, it is proposed that, given the qualitative nature of the data available on social media, the PRISMA guidelines can be re-purposed and applied to systematically collect and analyse this data.

3.2 Re-purposed PRISMA guidelines and checklist

Due to the nature of the data posted on social media, the four-phase PRISMA statement and 27-item checklist [95] were re-purposed and are described in Figure 27 and Table 8. The re-purposed PRISMA statement (Figure 27) is a flow diagram that depicts the flow of information through the four-phases of a systematic review; identification, screening, eligibility and inclusion, and maps out the number of records included and excluded during each phase. A social media data point or record, in this context, is defined as a single post. The re-purposed 27 checklist items (Table 8) pertain to the content of the systematic review of social media data and include the title, abstract, introduction, methods, results, discussion and funding associated with the study.

Figure 27. The re-purposed PRISMA statement for use in reviewing social media content



3.3 Discussion and implications for this PhD research project

Methods for extracting and analysing qualitative data from social media are manual and automated in nature, as was discussed in Chapter 2.2, and can present researchers with both benefits and limitations in their application. The number of studies that apply automated methods for extracting and analysing data is increasing (Figure 21). When this PhD research project was initiated in 2013, manual techniques were used in 42% of the eligible NCD studies without consistent guidance or application of methods.

The content posted on social media is often qualitative in nature and, in theory, can be subjected to systematic review methods, such as PRISMA. It is important, however, to acknowledge that the data used as input into the review process is fundamentally different. Traditional systematic reviews synthesise the results from multiple studies and can be considered a form of secondary research. A review of social media data undertaken systematically is a form of primary research where the social media data extracted and analysed are in their raw form.

The experience of conducting four empirical studies that analysed data from different social media (described in detail in Chapters 5-8), and of applying the PRISMA guidelines as part of the research design, has informed where these guidelines can be re-purposed and used effectively to review social media data, as well as where the applicability of this research method is limited (Table 8).

Table 8. The re-purposed checklist of items when reporting a review of social media content undertaken systematically

Section/Topic	**Item No.**	**Checklist item applied to data from social media**	**Applicability and limitations when used to review data from social media**
Title			
Title	1	Identify the report as a systematic review of social media with reference to the specific social media.	Applicable
Abstract			
Structured summary	2	Provide a structured summary including, as applicable, background, objectives, social media included, eligibility criteria for the data, synthesis methods, results, limitations, conclusions and implications of key findings.	Applicable
Introduction			
Rationale	3	Describe the rationale for the analysis of the social media data in the context of what is already known.	Applicable
Objectives	4	Provide an explicit statement of the questions being addressed.	Applicable
Methods			
Protocol and registration	5	Indicate if a review protocol exists and where it can be accessed.	Applicable
Eligibility criteria	6	Specify social media data point characteristics such as source, dates posted, language, terminology, and relevance.	Applicable
Information sources	7	Describe all social media included in the study.	Applicable
Search	8	Present the full method for extracting the data from the social media platform including any limits used, such that it could be repeated.	Applicable, although there may be limitations when considered in the context of other social media platforms and the accessibility of the data.
Study selection	9	State the process for selecting the relevant social media data from the selected social media (that is, screening, eligibility, included in systematic review and, if applicable, included in the meta-analysis).	Applicable due to the search criteria and pre-defined inclusion and exclusion criteria.
Data collection process	10	Not applicable to social media as already described in item number 8.	Not applicable to social media as already described in item number 8.
Data items	11	List and define all variables for which the social media data points were sought.	Applicable
Risk of bias in individual studies	12	Describes the methods used for assessing the risk of bias of individual data points and how this is to be used in any data synthesis.	The risk of bias in individual social media data points is dependent on the type of social media data included in the review. The data from social media is limited by character length and therefore the risk of bias within individual data points in this case is limited. This was not applied in the empirical studies (Chapters 5-8) and therefore limits the conclusions that can be drawn as to its applicability.
Summary of measures	13	State the principle summary measures.	Was not applied in the empirical studies (Chapters 5-8) and therefore this limits the conclusions that be drawn as to its applicability.
Synthesis of results	14	Describe the methods for handling the social media data and combining the results of multiple social media data points.	Applicable, although it would be dependent on the type of social media data under review (e.g. blog entries, videos) as to which method of synthesis is most relevant.
Risk of bias across studies	15	Specify any assessment of risk or bias that may affect the cumulative evidence from the social media data.	Applicable

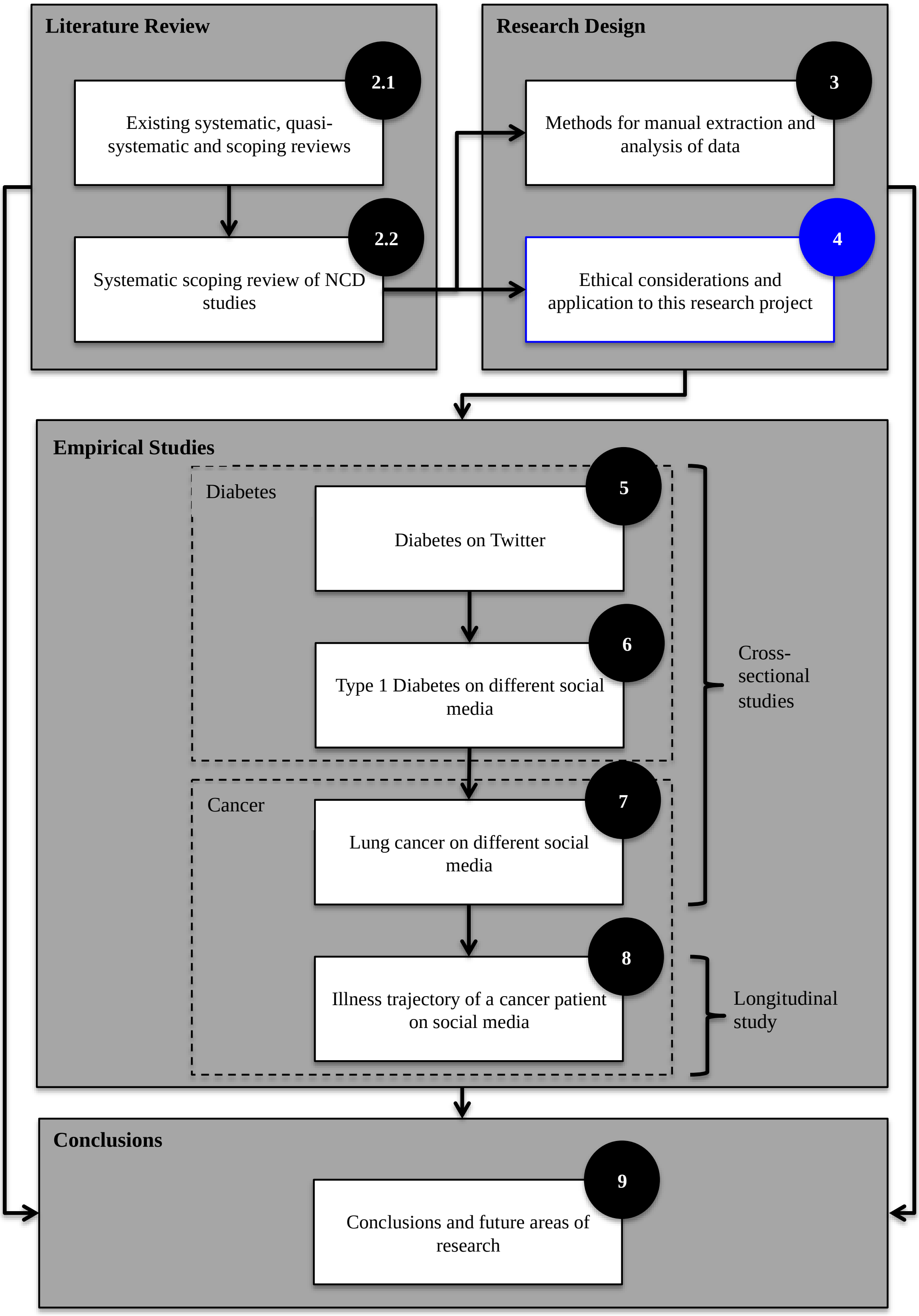
Section/Topic	Item No.	Checklist item applied to data from social media	Applicability and limitations when used to review data from social media
Additional analyses	16	Describe methods of additional analyses on the social media data.	Applicable
Results			
Study selection	17	Give the volume of social media data points screened, assessed for eligibility, and included in the study, with reasons for exclusion at each stage.	Applicable
Study characteristics	18	For each social media data point, present the characteristics for which the data was extracted (e.g. date stamp).	Applicable
Risk of bias within studies	19	Present the data on risk of bias of each data point and, if available, any outcome level assessment (see item 12).	The risk of bias within the social media data points is dependent on the type of social media data included in the review.
Results of individual studies	20	From all outcomes considered present, for each social media data point, a simple summary of the data and coding classification used.	Applicable to the classification of individual data points.
Synthesis of results	21	Present the results of each meta-analyses done, including confidence intervals and measures of consistency.	Dependent on the type of social media data under review. A formal meta-analysis was not applied in the empirical studies (Chapters 5-8) and therefore this limits the conclusions that can be drawn as to its applicability.
Risk of bias across studies	22	Present the results of any assessment of risk of bias across the social media data points.	Risk of bias across social media data points could be applied through a risk assessment of the social media data. This was not applied in the empirical studies (Chapters 5-8) and therefore limits the conclusions that be drawn as to its applicability.
Additional analysis	23	Give the results of additional analyses.	Applicable
Discussion			
Summary of evidence	24	Summarise the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (such as healthcare providers, users and policy makers).	Applicable
Limitations	25	Discuss limitations at the study and outcome level (such as risk of bias) and at review level (such as the social media data sources, extraction method, social media data points).	Applicable
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Applicable
Funding			
Funding	27	Describe the sources of funding for the study, other support and the role of funders.	Applicable

Although systematic review methodologies can be re-purposed and applied to the analysis of data available on social media, there are some limitations. These limitations include the use of PRISMA compared to other systematic and scoping review methods, of which Grant and Booth identified 14 different types [268], and other guidelines that are available, such as those published by Pickering and Byrne [269] and The Campbell Collaboration [270] to name but a few. The different unstructured forms of data posted on social media (e.g. videos, pictures, photographs), as well as the constantly evolving nature in which data is created, stored and used on social media are also limitations.

Compared to traditional systematic reviews, it can be argued that the risk of publication bias is potentially reduced or increased in the context of systematically reviewing posts on social media. This is due to the removal of barriers to publication and authorship, as anyone with access to the Internet and with a social media account can author content.

Applying critical appraisal and quality assessment techniques to the data posted on social media is more difficult to achieve, due to the unstructured nature and source of the data involved. The data posted on social media is raw data compared to the secondary data reviewed in traditional systematic reviews. Although account IDs for authors are available, it can be difficult to determine the provenance and motive behind the posts. Although qualitative methods for analysing data from social media were available, at the time of writing, there are no known techniques or guidelines available to critically appraise and assess the content and quality of the data posted on social media and therefore further research would be required to understand this topic and the implications for health research that analyses the data from social media.

Despite these limitations, the experience gained from applying the PRISMA guidelines to the manual analysis of social media data in the four empirical studies which form part of this PhD research project, and are described later in Chapters 5-8, concludes that, although the qualitative data posted on social media is primary data, many of the requirements for a systematic review can be applied. The experience of applying the re-purposed PRISMA guidelines to a corpus of gestational diabetes Tweets was presented and discussed at a poster presentation at the Medicine 2.0 Summit and World Congress in USA in 2014, details of which are provided in Appendix A.1.2.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

4. Ethical challenges and considerations when analysing data posted on social media

As well as the re-purposed systematic review method for manually analysing data from social media method (Chapter 3), the research design of any study involving data about individuals also requires ethical considerations to be taken into account. Reshnik [271] defines research ethics as the 'norms for conduct that distinguish between acceptable and unacceptable behavior', with its importance being a means to:

- Promote knowledge, truth and avoidance of falsification or misrepresentation of research.
- Promote fairness, trust and mutual respect essential for collaborative work.
- Ensure researchers are held accountable to the public for their conduct.
- Help build public support and funding for research that meets expectations of quality and integrity.
- Promote adherence to moral and social values of society such as human rights, compliance with the law and public health and safety.

The results of the systematic scoping review of the literature (Section 2.2, Figure 26) highlighted that when this PhD research project began in 2013, 45% (n=15 out of 33) of NCD-related studies made no reference to ethical considerations. Despite the increased number of NCD-related studies and the increased awareness and attention within the research community of ethical considerations when using innovative methods and data sources, 42% (n=44 out of 105) of the identified studies still do not reference the ethical considerations of sourcing data from social media. This continues to indicate a widespread neglect of ethical issues amongst research practitioners using social media data in NCD-related studies beyond the requirement for IRB/regulatory approval (41%, n=43).

This chapter further describes: 1) the ethical challenges faced by researchers, such as myself, who are conducting studies using data available on social media platforms, 2) examines how the Research Councils United Kingdom (RCUK) and affiliated research ethics guidelines acknowledge and deal with research involving social media overall, and specifically research involving data extracted from social media and 3) the implication of these findings on the conduct of the empirical studies of this PhD research project.

4.1 Challenges and the need for ethics

The complexity of interactions between individuals, groups and technical systems in online spaces, such as social media, presents a number of challenges for academics wishing to use social media data in research [272]. These include the self-selecting nature of users of these platforms, inequalities in access to these sites and the data they generate, the difficulty of obtaining meaning from heterogeneous data of variable quality and provenance and a dependence on observing and interpreting what is 'out there' in a way that differs from traditional sampling approaches. Arguably, however, the greatest challenges for researchers in this area are ethical ones [273, 274], such as variable perceptions of and unclear boundaries between 'public' and 'private' spaces, as well as the difficulty of ensuring anonymity and preserving the privacy of data subjects, whose identities may not be disguised or may be easily deduced from their postings and affiliations. Related issues of ownership and intellectual property are also poorly defined and consent to the use of social media data in research is rarely obtained through informed choice, but rather assumed on the basis that users have chosen to place it in the public domain [272, 275-279]. Awareness of the potential privacy implications of sharing personal information on social media is growing, driven by newsworthy cases such as Facebook's experiments in emotion manipulation [280] and its identification of 'vulnerable' teenagers for advertisers [281] or the use of social media by data analytics companies seeking insights into citizens' political attitudes and networks to influence voter behaviour [282, 283]. In this environment, pinning down the ethical guidance for researchers is now more critical than ever, with a requirement for any guidance to be responsive and adaptable to the changes invoked by the rapid evolution of social media platforms and data science.

Most research institutions, irrespective of academic discipline, publish or adhere to some form of research ethics guidelines or standard operating procedures, as a means of ensuring the appropriate governance of studies undertaken by their staff and collaborators. This is a standard requirement for any UK Higher Education Institution receiving Department of Health research funding. While these vary in structure, content and application, they are all intended to ensure responsible and trustworthy research practice and 'to protect all groups involved in research: participants, institutions, funders and researchers throughout the lifetime of the research and into the dissemination process' [284:pg2]. As discussed in earlier chapters, social media research is still a relatively new and changing field and commentators have pointed to the destabilisation of traditional ethics and an unsettling of ethical expectations and assumptions for both researchers and Internet users [285]. This has been compounded by a lack of internet research ethics guidance at both national and international levels [262], the deficit of which poses particular challenges for research involving ‘sensitive’ data, such as information about vulnerable groups e.g. children, people in care, people’s health conditions, political affiliations or religious beliefs [286].

4.2 How the RCUK ethics guidelines acknowledge and address research involving social media overall and, specifically, research using data extracted from social media

Given the growth of research using social media platforms, as described in Chapter 2.2, and its potential implications for information privacy, confidentiality and ownership, a study was conducted as part of this PhD research project to examine the extent to which existing research ethics guidelines take such uses into account and what additions may be warranted. Social media research is taking place across multiple academic disciplines and applications for research ethics approval may thus refer to a range of different bodies. This presents challenges for the effective oversight of such research where, it has been claimed, ‘no official guidance or answers regarding internet research ethics have been adopted at any national or international level’ [262:pg 2]. Mindful of the need for a cross-disciplinary perspective, one identifiable national corpus of multi-disciplinary research ethics guidelines was chosen, represented by Research Councils United Kingdom (RCUK).

RCUK is a strategic partnership between the UK’s seven research councils (Box 4) which, according to its homepage, ‘has invested around £3 billion in research covering the full spectrum of academic disciplines from the medical and biological sciences to astronomy, physics, chemistry and engineering, social sciences, economics, environmental sciences and the arts and humanities’ [287]. The partnership shares the aim to ‘advance knowledge and generate new ideas which lead to a productive economy, healthy society and contribute to a sustainable world’ [287]. While RCUK itself has published a set of general research ethics guidelines, each of the seven disciplinary bodies in the RCUK family of research councils provides its own form of ethical advice, either through developing bespoke guidelines or by deferring to other relevant guidelines in the literature. For the purposes of this study, the corpus of RCUK ethics guidelines and external guidelines recommended within these was felt to be an appropriate sample to enable a meaningful analysis of the guidance available for academic researchers in the UK.

Box 4. RCUK umbrella organisation and the seven UK Research Councils

- Research Councils United Kingdom [288]
- Arts and Humanities Research Council [289]
- Biotechnology and Biological Sciences Research Council [290]
- Engineering and Physical Sciences Research Council [291]
- Economic and Social Research Council [292]
- Medical Research Council [293]
- Natural Environment Research Council [294]
- Science and Technology Facilities Council [295]

4.3 Methods

To aid the analysis, inspiration was drawn from two ethical frameworks described earlier in this thesis which, although developed in the context of social media research for health, are sufficiently generic to be applied to any field of research involving the use of social media.

The first is Bjerglund-Andersen and Söderqvist's [25] typology of social media uses in research, which delineates five broad categories:

1. Research dissemination;
2. Scientific discussion and networking;
3. Engaging the public;
4. Academic teaching; and
5. Research and data collection.

For the reasons already described, the last of these was divided into two qualitatively different categories – first, using social media platforms to *enable the conduct of research* and, secondly, using social media as a *source of data for research.*

Conway [99] has gone further, by suggesting a taxonomy of ethical considerations specifically relevant to the secondary use of social media data. Although this was developed in the context of Twitter mining for public health surveillance and research, it is applicable to many types of research involving data sourced from social media. This includes 10 specific considerations:

1. Privacy;
2. Informed consent;
3. Ethical theory;
4. IRB/regulation;
5. Traditional research versus social media (e.g. Twitter) research;
6. Geographical information;
7. Researcher lurking;
8. Economic value of personal information;
9. Medical exceptionalism; and
10. Benefit of identifying socially harmful medical conditions.

Although considerations 9 and 10 were developed in the context of medical research, they can also be applied to other topics which are also uniquely sensitive (e.g. research on political attitudes) or are aimed at preventing harm (e.g. analysing extremist discourse), respectively. For the broader purposes of this study, these considerations were therefore re-labelled them as ‘exceptionalism’ and ‘benefit of identifying potential harms’.

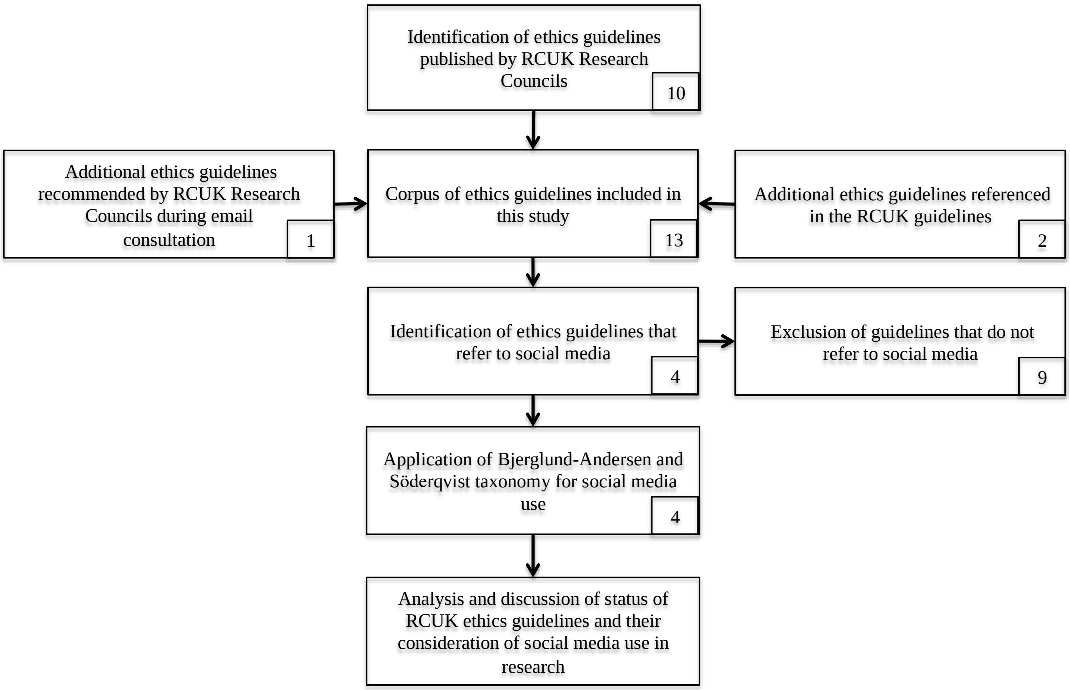
To identify the corpus of ethics guidelines represented by RCUK, the websites of RCUK itself and the seven UK Research Councils were first identified via Google. The websites were then searched by entering the key words ‘ethics’, ‘guidelines’, ’funding applications’, and variants of these into their respective search boxes and the outputs sifted manually. Searches were undertaken in February 2017.

Where a research agency was found to have more than one current ethics guideline, each of these was included and, in cases where the RCUK guidelines explicitly referred to external guidelines, the relevant source documents were also obtained for further analysis. Individual research councils were also contacted via email, asking them to state whether their organisation had developed or specifically recommended any ethics guidelines concerning the use of social media in research. Responses were received from six out of eight agencies, the non-respondents being the BBSRC and the ESRC.

The following information was extracted from each identified guideline: the name of the originating organisation, the title of the guideline, the date of the most recent version and whether the guideline explicitly referred to the use of social media or related concepts, such as online or internet research.

The four guidelines referring to social media were scrutinised to determine how they corresponded with the (adapted) typology of social media uses in research outlined by Bjerglund-Andersen and Söderqvist. They were further appraised in terms of their reference to Conway's list of 10 ethical considerations for research involving social media data. The guideline search and appraisal process is summarised in Figure 28.

Figure 28. Summary of the search and appraisal process for the ethical guidelines



4.4 Results

Thirteen separate sets of ethics guidelines were identified, including ten produced by RCUK itself or the individual UK research councils, two external guidelines recommended within these [261, 262] and one recommended by MRC during the email verification phase [20]. Of these, only four guidelines (ESRC, BPS, AoIR, NIHR) mentioned the use of social media in research.

The 13 guidelines are listed in Table 9, which also illustrates the co-referencing of guidelines within the RCUK family; for example, ARHC’s guideline refers to the ESRC’s guideline which, in turn, cites guidelines from the BPS and AoIR. Highlighted in bold are the four guidelines found to include guidance and recommendations specifically relating to the use of social media in research such as those published by ESRC, BPS, AoIR and BPS.

Table 9 illustrates a further level of analysis, focused on the four guidelines that encompassed social media. Based on the adapted version of Bjerglund-Andersen and Söderqvist’s taxonomy, all four referred to social media as a research tool, three as a source of research data, two each as a medium for scientific discussion, networking or public engagement and none for research dissemination or academic teaching. According to Conway’s list of ethical considerations in social media research, all four of these guidelines referred to privacy and the difference between traditional and social media research, three referred to informed consent and the use of IRBs, two referred to researcher lurking and one to ethical theory. None considered geographical information, the economic value of personal information, exceptionalism or the benefit of identifying sources of potential harm.

Table 9. Ethics guidelines screened for references to social media uses in research

Research Council (date)	**Guideline title**	**Includes**	**Refers to social media**
RCUK [296]	Policy and Guidelines on Governance of Good Research Conduct		
AHRC [297]	Research Funding Guide	RCUK ESRC	
BBSRC [298]	BBSRC Research Grants The Guide		
EPSRC [299]	Framework for Responsible Innovation	RCUK	
MRC [300]	Policy and Guidance on Sharing of Research Data from Population and Patient Studies		
MRC [301]	Personal Information in Medical Research		
MRC [302]	Good research practice		
NERC [303]	Ethics Policy		
STFC [304]	Public Engagement with Science and Technology	MRC RCUK	
ESRC [305]	**Framework for research ethics**	BPS AoIR RCUK	**x**
BPS [261] ***In ESRC***	**Guidelines for ethical practice in psychological research online**		**x**
AoIR [262] ***In ESRC***	**Ethical Decision Making and Internet Research**		**x**
NIHR [20] ***Recomm MRC***	**Guidance on the use of social media to actively involve people in research**		**x**
TOTAL			**4**

Table 10. Types of research use and ethical considerations for data re-use

Research Council	**Guideline Title**	**Types of social media use in research**						**Ethical considerations for the use of social media data in research**									
		1	**2**	**3**	**4**	**5**	**6**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
ESRC	Framework for research ethics					x	x	x	x	x	x	x					
BPS (In ESRC)	Guidelines for ethical practice in psychological research online			x		x	x	x	x			x		x			
AoIR *(In ESRC)*	Ethical Decision Making and Internet Research		x			x	x	x	x		x	x		x			
NIHR *(Recomm. MRC)*	Guidance on the use of social media to actively involve people in research		x	x		x		x			x	x					
Total		0	2	2	0	4	3	4	3	1	3	4	0	2	0	0	0

Key for Table 10:
Bjerglund-Andersen & Söderqvist's classes of social media use (adapted): 1=Research Dissemination, 2=Scientific discussion/networking, 3=Engaging the public, 4=Academic teaching, 5=Social media as a research tool, 6=Social media as a source of research data.

Conway's ethical considerations for social media data use: A=Privacy, B=Informed consent, C=Ethical theory, D=IRB approval/regulations, E=Traditional vs social media research, F=Geographical information, G=Research lurking, H=Economic value of personal information, I=Exceptionalism, J=Benefit of identifying sources of potential harm.

4.5 Discussion

The analysis conducted as part of the systematic scoping review of the literature (Section 2.2, Figure 26) highlighted that 42% (n=44) of NCD-related studies made no reference to ethical considerations. When combined with the examination of research ethics guidelines (Section 4.2), this suggests gaps in the application of such guidelines in research using data mined from social media, as well as an incompleteness and inconsistency of current guidelines and an absence of ethical discourse in published research articles.

Status of RCUK ethics guidelines on social media and social media data

Of the seven multi-disciplinary ethics guidelines published by RCUK, only one (published by ESRC) specifically considered the use of social media in research, despite such research now straddling the remits of many national funding agencies. Two research councils (ESRC, MRC) nevertheless recommended guidelines from other bodies (AoIR, BPS, NIHR/INVOLVE), generating a corpus of four social-media-relevant guidelines for UK researchers. These referred to social media as a research tool (4/4), as a source of data (3/4), as a means of public engagement (2/4) and as a channel for scientific discussion and networking (2/4), but did not mention their use for research dissemination or teaching, which also appear in the adapted version of Bjerglund-Andersen and Söderqvist’s taxonomy. With specific reference to the mining and re-use of social media data, these guidelines prioritised privacy (4/4), differences between digital and conventional research (4/4), informed consent (3/4), IRB approval/regulation (3/4) and researcher lurking (2/4), although none of the other four ethical considerations in Conway’s (2014) framework were covered. In Section 2.2 of this thesis, medicine was highlighted as the dominant subject area of authors (Figure 12) publishing NCD-related research that sources data from social media. Although MRC was the source of three research ethics guidelines, none referred to the use of social media, in contrast to their detailed consideration of ethical issues surrounding the re-use of institutional and research datasets, where most of the UK’s ‘big data’ investments are taking place. ESRC provided the most comprehensive overview of social media ethics, also deferring to the external AoIR and BPS guidelines, likely reflecting the importance of digital social research within ESRC’s portfolio. While these differences between research councils are, to some extent, understandable given that the BBSRC and NERC may be less inclined to fund research involving social media due to its limited relevance to them, they indicate a segmentation of data ethics along disciplinary lines, which is unhelpful in an environment where multi-disciplinary projects are the norm, rather than the exception, underscoring the need for collaboration and agreement on universal principles.

As already noted, ESRC was the only UK research council to provide individual ethics guidelines explicitly considering the use of social media in research. Its *Framework for Ethics* [305] includes a detailed overview of relevant issues, along with examples, and illustrates the potential for ethics guidelines to evolve in response to emerging innovations. While the earlier version of this framework [284] advised that research involving respondents through the Internet may '*involve more than minimal risk*', no specific examples of risk were provided to guide researchers in this assessment. This lack of specific guidance was also reflected in the 'frequently asked questions' section dealing with Internet searches, where it was simply noted that the rapidly evolving nature of the field and the use of web pages and instant messaging for research purposes '*pose new ethical dilemmas*' that need to be addressed. In contrast, the guidelines published in January 2015 refer explicitly to ethical considerations associated with the use of social media as a research tool and as a source of research data. These include uncertainties over how to apply ethical concepts such as 'privacy' and 'anonymity', which may be interpreted differently by social media users and researchers, and the potential sensitivity of topics discussed in these settings, such as health issues. They caution that, while information intentionally published on the Internet is 'in the public domain', the identity of individuals should be protected unless it is critical to the research, such as in studies analysing statements by public officials. ESRC's 2015 guidelines also advise researchers to abide by the regulations and permissions set by the data holders (e.g. Twitter, Facebook), particularly when these are required for compliance with data protection legislation, bearing in mind that such research may cross legislative jurisdictions. The framework also benefits from deferring to two internet-specific research ethics guidelines developed by the BPS and the AoIR.

While the BPS guidelines, published in 2012, do not explicitly refer to social media, this is implied in the term 'internet-mediated research', which encompasses the use of online platforms as means of engaging the public, as a set of research tools and as a source of data for secondary uses, consistent with our taxonomy. Ten ethical considerations are highlighted, which overlap with, but are somewhat different from, those provided by Conway. These include verifying identity, private versus public space, informed consent, levels of control, withdrawal, debriefing, deception, monitoring, protection of participants and researchers and data protection. These are grouped into four sectors of a grid, relating to whether participants are actively recruited or are unaware of their involvement in the study, as well as whether they are identified or anonymous. Although the BPS guidelines go some way towards providing actionable recommendations for researchers, they should not be considered exhaustive, given that only four of the ten ethical concepts identified by Conway (privacy, informed consent, IRBs and researcher lurking) are addressed. A newer BPS guideline, which was under beta-testing at the time of this study, extended on the 2012 framework but, as yet, does not refer to social media specifically (BPS, 2017). Based on this study, it was recommended that this be included. The official guidelines published by BPS in 2017 [306] did refer to social media explicitly in the context of four main principles underpinning the ethical conduct of research: 1) respect for the autonomy, privacy and dignity of individuals and communities, 2) scientific integrity, 3) social responsibility and 4) maximising benefits and minimising harm. These guidelines, however, still only take into consideration six of the ten ethical concepts identified by Conway.

The AoIR is a widely recognised international academic association dedicated to the advancement of the cross-disciplinary field of Internet studies. The AoIR ethics guideline referred to by the ESRC [262], was published in 2012 and outlines several high-level themes, including the difficulty of understanding whether such research involves ‘human subjects’ for the purposes of ethics approval, differentiating ‘public from private’, conceptualising data or text as an extension of ‘persons’ and reconciling ‘top down versus bottom-up approaches’ for managing potential harms and benefits of research. The document includes an extensive list of considerations, such as understanding the context of the research, the primary objective of the research, how the data will be accessed, stored and disseminated and the rights of participants, who may be unaware that their data are being used. Unlike the 2012 BPS guideline, the AoIR guidelines of 2012, explicitly mention social media and gives examples of social media data uses that present ethical challenges. The AoIR later published Version 3.0 of their ethical guidelines in 2019 [222] and provided further guidance to IRBs and researchers that included a taxonomy of different stages of research (i.e. initial research design, initial research processes, analyses, dissemination of the research and the close of the project) as well as a process for ethical analysis at each stage.

Given the potential sensitivity of medical information available online and the findings from the systematic scoping review (Section 2.2) that highlight the volume of NCD-related studies being published by researchers affiliated a medical subject area, it is somewhat surprising that the MRC does not provide specific guidance for researchers conducting studies using social media data. Nevertheless, in its email verifying this, the MRC recommended that the guidance provided by the NIHR as part of the INVOLVE advisory group be reviewed. INVOLVE was established by NIHR in 1996 to support active public involvement in NHS, public health and social care research. In 2014, it published ethics guidelines on using social media to engage citizens in public debate and research, as a forum for scientific discussion and networking and as tool for undertaking research and consultation. It lists the types of social media available, provides case studies of their use, outlines the benefits and challenges, considers how to manage risk and offers tips based on researcher experience. Applying Conway's taxonomy to the NIHR guidance, however, indicates that only three of the 10 ethical concepts are addressed, namely: privacy, the use of IRBs and the difference between traditional and social media-based research. These reflect the public-engagement remit of INVOLVE, which may explain why the secondary use of social media data for research is not explicitly discussed.

The absence of any reference to research using social media in the remaining RCUK guidelines is noteworthy. Whilst in some cases this is entirely understandable – for example, the STFC focuses primarily on particle and nuclear physics and science infrastructure – in others it would seem appropriate to include these new forms of data. For example, one EPSRC project specifically focuses on the use of social media, crowdsourcing and citizen science, albeit driven by computer scientists [307]. This project includes themes in health and social science**,** illustrating how social media research transects disciplinary boundaries and may potentially fall within the scope of several ethics bodies.

The following quotation from the AoIR [262] guideline neatly illustrates the need for this trans-disciplinary thinking.

'Manipulation and close study of information generated by social media networks certainly constitutes a different research environment than sticking a needle into a volunteering person in a medical laboratory. On the other hand, entire communities have felt harm from use of their DNA data more than a decade after it was collected and anonymously aggregated.' pg.13

Policy implications

Despite their use now being common, the emergence of social media and other online platforms has taken traditionally slow-moving governments and academic institutions somewhat off-guard. Uncertainties about what is appropriate, acceptable, legal and responsible in these new virtual spaces, and for different forms of digital personal information, has also fuelled broader debates. These include debates around the need for ‘net neutrality’ or equal access to internet content and services amongst all users [308], how to maintain control of key Internet domain names in the global public interest [309] and calls for a ‘Magna Carta for Data’ [310, 311]. Moreover, it is contributing to the dilemma of governments seeking to generate economic, scientific and societal value from existing data assets, whilst also protecting citizens from unwanted surveillance and intrusion. Health research is one area in which this discussion has been particularly acute, due to the traditionally stringent ethical demands placed on the protection of confidentiality. In the UK, the growing use of health records for research [312], coupled with public disquiet over controversial programmes such as Care.Data [313] and Google DeepMind’s Streams project [314], has focused considerable policy attention on the need for ethical and robust governance when it comes to the use of patient information (e.g. Richards et al, 2015; National Data Guardian [315]). In this context, it is noteworthy that, by comparison, the ethics of using social media data in health research has been somewhat neglected, albeit such data is seldom managed by the state or by healthcare institutions with a duty to protect it. It is nevertheless arguable that the same principles of respect, confidentiality and protection from harm or embarrassment should be followed, as would be expected in any other form of *bona fide* research.

Caveats and opportunities for further research and development

The review of ethics guidelines was limited to those provided or recommended by RCUK and its seven UK Research Councils, as well as the other relevant guidelines developed by UK-based researchers [316] and organisations beyond the scope of this study [e.g. 317]. It is recommended that further research involving a wider corpus of research ethics guidelines be conducted, to test the generalisability of these results in the UK, and as a means of catalysing the development of internationally-applicable ethics guidelines for research involving social media platforms and data.

The variable coherence, consistency and navigability of the RCUK websites presented a challenge in terms of identifying relevant ethics guidelines, particularly in the case of MRC and EPSRC. For MRC, this was mainly due to its diverse portfolio of specialised guidelines, covering topics ranging from clinical trial management to the use of human tissue samples. For EPSRC, the distribution and annotation of ethical information represented more difficulties, with a list of high-level ethical considerations accompanied by hyperlinks to the RCUK framework and a variety of external sources, many with little or no annotation. One exception is the ‘*Framework for Ethical and Responsible Innovation’*, which arose from an EPSRC-funded research project and is referenced repeatedly on the website, although its full text is only accessible via a hyperlink to the authors' journal proof. It is recommended that action to improve consistency amongst RCUK members in their presentation of ethical guidance, including appropriate content tagging, be undertaken to avoid confusion and facilitate access to relevant advice for researchers using social media in their studies.

Following the Alder Hey Hospital scandal in 1999, the Department of Health mandated that Higher Education Institutions receiving funding be required to have a research ethics policy [318]. The multiplicity of departmental and institutional ethics committees operating within UK universities and research organisations adds further complexity to the landscape of ethical governance and practice. New empirical studies are needed to shed light on the ways in which such committees are addressing approval requests for studies involving the reuse of data from social media, including which published guidelines they refer to, whether they have their own written policies for this type of research and whether disciplinary affiliation affects decision making.

The scope of this analysis did not extend to legal or regulatory aspects of information governance in the context of social media data, which are designed to control or limit certain forms of research. In contrast, ethical guidelines aim to ensure research integrity, discourage irresponsible or socially unacceptable research conduct and support the prioritisation of studies likely to benefit rather than harm society. Likewise, a comparison was not sought of methodological innovations such as automated data mining methods, social network analysis, machine learning and the processing of personal data by 'black box' algorithms, which also present challenges around consumer choice, control and privacy [319]. Comparable analyses conducted from each of these perspectives are warranted.

4.6 Conclusions

The results of this examination into how research sponsors and researchers are addressing the ethical challenges of analysing data posted on social media were published in the *Research Ethics* journal in 2017, details of which are provided in the Appendix (A.2.2).

Awareness and variability of guidelines

Beyond statements about IRB approval, the generally poor integration of ethical concepts and guidelines within the corpus of published articles reviewed in Section 2.2, suggests low levels of awareness amongst researchers using social media mining in their studies, echoing observations from other areas of ‘big data’ research [320]. This is consistent with the wide variability observed in the research ethics guidance offered by RCUK members in relation to uses of social media platforms and the data derived from them. BBSRC and NERC are potentially less likely to fund research involving social media compared to the ESRC, however there is a clear need for more detailed guidance from research councils such as the ESRC and EPSRC when conducting studies that involve large scale data processing from social media within computer science. Findings from this study indicate that only one RCUK council (published by ESRC) directly refers to social media research in its ethical guidance is a cause for concern, given the highly inter-disciplinary nature of studies in this area.

Consistency and accessibility of guidelines

It is recommended that further cross-council collaboration is needed to develop shared, inter-disciplinary, guidelines for the ethical use of social media in research, and specifically research involving the harvesting and reuse of social media data. In the shorter term, effort should be invested to improve consistency in the presentation, accessibility and comprehensiveness of existing ethical guidance available on the various RCUK websites. For example, it was observed that some websites are difficult to navigate and contain highly distributed and poorly connected information on ethics, approval processes and regulation. Adequate literature review to ensure the timely inclusion of relevant guidance from other sources is also required; for example, a guide to ethics in social media research which had emerged from a project part-funded by ESRC and EPSRC, was identified but was not mentioned on either of their websites [321].

Future RCUK ethics guidelines would also benefit from including a broader range of social media uses, clear criteria for judging projects against a variety of ethical considerations, and pragmatic recommendations for researchers planning to undertake studies involving social media. Until such meta-guidelines are available, the recommendation was that, until such meta-guidelines are available, UK researchers prioritise the existing guidelines produced by the ESRC, BPS, AoIR and NIHR, alongside the adapted ethical taxonomy provided earlier in this chapter. Researchers are also encouraged to explore the wider universe of ethical frameworks emerging nationally and internationally in relation to new forms of data, including those from the OECD [322], the US Council for Big Data Ethics and Society [320] and the UK Data Service [323] as well as emerging initiatives such as the UK Society for Data Miners’ plans to develop ethical principles [324] and primary research exploring the boundaries of public acceptability in the reuse of digital personal data [325, 326].

The role of authors, editors and funding bodies

It was also recommended that UK researchers applying for project funding or permission to undertake studies using social media data should explicitly state which ethics guidelines they have consulted, and that IRBs integrate this requirement into their approvals documentation. A call is made upon authors and editors to ensure that publications describing studies involving social media data clearly state the ethical issues that have been considered during the research and specify the guidelines consulted.

Given the substantial investments made in digital research and data science by the UK government and research councils over the last five years, coupled with increased policy attention on responsible research and innovation [327] and the protection of personal data by means of the European General Data Protection Regulation (GDPR) introduced in 2016 [328], ensuring the robust design and implementation of ethical guidelines for social media research is essential.

It is hoped that the results of this scoping study of ethical guidelines will inform the future development of such guidelines in the UK and elsewhere, and catalyse a broader inter-disciplinary discussion amongst research councils, institutional ethics boards and researchers themselves.

4.7 Implications for this PhD research project

Based on these recommendations and in preparation for conducting the four empirical studies described in Chapters 5-8 of this thesis, Conway's adapted ethical taxonomy was reviewed and six of the ten ethical concepts were considered relevant to the studies. These are, namely, privacy, informed consent, ethical theory, IRB/regulation, traditional vs social media research and researcher lurking. The ethical considerations and guidelines provided by ESRC, BPS, AoIR and NIHR and their application to these studies are described in Tables 11, 12, 13 and 14, respectively.

Table 11. Application of the ESRC 'Framework of Research Ethics'

Section	ESRC guideline / recommendation	Application to this PhD research project
Our principles and expectations for ethical research (pg. 5)	Research involving primary data collection will always raise issues of ethics that must be addressed. Whilst the re-use of some datasets may be relatively uncontroversial and require only light-touch ethics review, novel use of existing data and especially data linkage, as well as some uses of administrative, internet-mediated data and controlled data, will raise ethics issues.	The ethical issues and guidelines associated with the secondary use of social media data for health research has been discussed in Chapter 4. These recommendations were applied to the empirical studies discussed in Chapters 5-8 of this thesis.
Research potentially requiring a full ethics review (pg. 10)	Social media and participants recruited or identified through the internet, in particular when the understanding of privacy in these settings is contentious where sensitive issues are discussed - for example in 'closed' discussion groups where there is potential for quotes to be identifiable and including where visual images are used.	Existing literature identifies theoretical risks associated with the secondary use of data from social media as; the variable perceptions of 'public' and 'private' spaces, difficulties preserving the privacy of data subjects, consent to the use of social media data in research and experimental interventions that influence emotions, attitudes and behaviours of data subjects without their knowledge. Measures taken to manage such risks during the conduct of this research project are identified as: • Only posts freely published on public social media were considered in scope. • Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than 10, i.e. Kate Granger's widower, Chris Pointon. • Data was aggregated to mitigate the risk of identifying individuals and no personal information was included in any associated publications. The exception to this was Kate Granger's widower, where consent was granted. • Direct contact with data subjects and experimental interventions was considered outside the scope for these observational studies.
1.4 RECs should be constituted and operate in accordance with the framework standards (pg. 12)	Information provided in forums or spaces on the internet that are intentionally public would be considered 'in the public domain', but the public nature of any communication or information on the internet or through social media should always be critically examined, and the identity of individuals protected, wherever possible, unless it is critical to the research, such as in statements by public officials. For research that involves the use of social media, researchers will also need to abide by the regulations set by the data producers subject to such regulations being consistent with legal and ethical guidelines (see also internet-mediated research in FAQs). The potential for identifiability of online sources, as well as ethical debates about how privacy is constituted in digital contexts, means that full ethics review may be appropriate for research involving these communities. For example, people often assume that social media sources are public domain, but it is quite likely that some service users – including children – may not understand the implications of what they are doing and those harvesting data may also uncover illegal images or	

Section	ESRC guideline / recommendation	Application to this PhD research project
	activities.	
Data access through technology (pg. 26)	Researchers are now making greater use of datasets, which have been generated through internet-mediated technology and social media. Researchers need to consider the ethical issues which arise; for example, the interpretation of anonymity and whether participants (e.g. social media users) would consider data in the public domain to be private, the meaning of informed consent in this context and the important issue of what permissions a researcher has over the data supplied by the data producer (eg Facebook or Twitter data).	
2.4 Internet-mediated research Why should internet research receive full ethics review? (pg. 34)	In a fast-developing area, RECs may need to involve an independent expert in assessing research proposals that break new ground. Internet research can take place in a range of settings, for example email, chat rooms, web pages, social media and various forms of 'instant messaging'. These can pose specific ethical dilemmas. For example, what constitutes 'privacy' in an online environment? How easy is it to get informed consent from the participants in the community being researched? What does informed consent entail in that context? How certain is the researcher that they can establish the 'real' identity of the participants? When is deception or covert observation justifiable? How are issues of identifiability addressed? Researchers, research participants and reviewers of research ethics will often encounter new or unfamiliar ethics questions and dilemmas. There is a growing literature on ethics in online research. A good starting point is the Association of Internet Researchers 2012, report and the BPS 'Ethics Guidelines for Internet-mediated Research' 2013.	These empirical studies (Chapters 5-7) intended to aggregate data from multiple hundreds of social media users in order to identify potential trends and population insights. Many users of social media do not use a pseudonym and therefore the raw data may still be identifiable. In order to preserve the anonymity of content authors, the data was aggregated and their account ID as well as any reference to other account IDs (e.g. the use of @) were removed. For Twitter, data was collected using hashtags on the open public platform. For closed Facebook groups, permission to access data was sought through contacting the group administrators and explaining the nature of the request and activity. None of the raw data collected was disseminated beyond those involved in the conduct of the study. Ethics approval was sought and given by the University of Edinburgh's IRB. Additional ethics guidelines produced by the BPS (Table 12) and AoIR (Table 13) were also applied.

Section	ESRC guideline / recommendation	Application to this PhD research project
Appendix A: Example research ethics initial checklist (pg. 38-39)	Have you considered risks to: - the research team - the participants, e.g. harm, deception, impact of outcomes - the data collected e.g. storage, considerations of privacy, quality - the ROs, collaborators, project partners and funders involved to anyone else be put at risk as a consequence of this research? What might these risks be? How can these risks be addressed? How will you protect your data at the research site and away from the research site?	Risk to research team was minimal and was mainly reputational in nature. By joining a closed Facebook group using a personal profile, researchers may divulge personal information about themselves to other group members. Privacy risks to individuals whose social media data were extracted and analysed for the purposes of the empirical studies (Chapters 5-7) was mitigated through aggregation of data, and removal of account IDs. The development of an online intervention based on the data collected was not in scope for the empirical, observational studies included in this research project.
	Details and recruitment of participants: - What types of people will be recruited? E.g. students, children, people with learning disabilities, the elderly - How will the competence of participants to give informed consent be determined? - How, where, and by whom participants will be identified, approached, and recruited? - Will any unequal relationships exist between anyone involved in the recruitment and the potential participants? - Are there any benefits to participants? - Is there a need for participants to be de-briefed? By whom?	No individual participants were actively recruited for the empirical studies. (Chapters 5-8) included in this research project. The self-selection of participants was due to the dependency on having a social media account in order to post a message. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten, i.e. Kate Granger’s widower, Chris Pointon.
	What information will participants be given about the research? Who will benefit from this research? Have you considered anonymity and confidentiality? How will you store your collected data? How will data be disposed of and after how long?	Those asked to provide consent (e.g. administrators of closed online community and widower of deceased), were provided information about the purpose and intent of the study. Those that benefit from this research include: The research community, through contributing to a growing body of knowledge that can inform future research

Section	ESRC guideline / recommendation	Application to this PhD research project
	Are there any conflicts of interest in undertaking this research, e.g financial reward for outcomes? Will you be collecting information through a third party?	and interventions. People affected by the condition, and those supporting them (e.g. healthcare teams, family, carers, clergy), who should consider online communities as an additional source of social support during times of crisis. Design of innovative palliative care services. No feedback was provided directly to participants, although aggregated findings were made available through publication in peer-reviewed, scientific journals. The research team conducting the secondary analysis of the data from social media had access to the sample of data extracted for the purposes of the study. Platform owners (e.g. Facebook, Twitter) and other social media users and members of these online communities have direct access to the raw data through the social media platform itself. Most of the extraction of data from social media for the empirical studies (Chapters 5-8) was done manually. The exceptions to this are: the use of an own script for the pilot study (Chapter 5) to identify the diabetes tweets and a third-party tool provided by Symplur to extract the sample of tweets using hashtags for the Lung Cancer study (Chapter 7). During the conduct of the empirical studies (Chapters 5-8), the data files were stored on a local hard drive and Dropbox. Data files were shared amongst the research team via the university email system and disposed of by ‘shredding’ them electronically, following successful publication of the research study in a scientific journal.
	Have you considered consent? If using secondary data, does the consent from the primary data cover further analysis?	Only posts freely published on public social media were considered in scope. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the

Section	ESRC guideline / recommendation	Application to this PhD research project
	Can participants opt out? Does your information sheet (or equivalent) contain all the information participants need? If your research changes, how will consent be renegotiated? Have you considered ethics within your plans for dissemination/impact?	sample of individuals included in the study was less than ten, i.e. Kate Granger's widower, Chris Pointon.
	Are you conducting research outside the UK? Are there any additional issues that need to be considered as a result? E.g. local customs, local 'gatekeepers', political sensitivities? Which Ethics Committee is most appropriate for your research? Have you considered the time you need to gain ethics review? Have you considered what legal requirements your project will need to abide by? E.g. Data Protection Act, Freedom of Information Act, Human Rights Act How will the ethics aspects of the project be monitored throughout its course? Is there an approved research ethics protocol that would be appropriate to use? How will unforeseen or adverse events in the course of research be managed? E.g. do you have procedures to deal with any disclosures from vulnerable participants? Have you considered data management and curation? What measures have been taken to ensure confidentiality, privacy and data protection during and beyond the end of the project and to encourage data sharing and linkage? See the ESRC Research Data Policy.	The geographical and political boundaries that exist for other forms of health research were not always evident in research that involves the secondary use of social media data. For the empirical studies described in this thesis (Chapters 5-8), English language posts were considered in scope. The empirical studies were conducted in line with the terms and conditions of use for each of the social media platforms. Ethics approval was sought and given by the University of Edinburgh's IRB. An ethics form was submitted in 2014 that focused on diabetes. This was later expanded in 2016 to include the identification, collection and analysis of online posts for other NCDs (e.g. cancer) using similar methods.

Table 12. Application of the BPS 'Guidelines for ethical practice in psychological research online'

Section	BPS guideline / recommendation	Application to this PhD research project
1. Verifying identity (pg. 2)	An obvious difference in psychological research conducted online is that the participants and researchers are not physically brought together when taking part in the research. It is therefore difficult to verify the identity of research participants. The potential impact of this issue will be related to the design of the research and the degree to which the nature of the study makes it critical to confirm the actual identity of the participant, or particular participant characteristics (such as gender or age). Researchers may employ a variety of mechanisms in an attempt to authenticate participant identity such as checks made by credit cards, banks, using passwords, or registration of an e-mail, user name and/or password. However, whilst these mechanisms may diminish the extent of fake identity, they are not failsafe and necessarily involve additional costs and time, as well as generally lowering the number of participants willing to take part. Researchers need to be aware of the difficulties in verifying identities online in the design of the research and the evaluation of risk and ethical practice. If it is essential to know the exact identity, or to constrain the range of participants taking part (for instance to exclude vulnerable or underage individuals), then Internet Mediated Research (IMR) may not be appropriate. The extent to which one can verify identity depends on the mode of access used. Later in these guidelines, the differing degrees of control over the authentication process are discussed.	Verification of identity is a known limitation of conducting research that involves data from social media. In order to mitigate against this risk and its potential impact on the validity of results from the empirical studies (Chapters 5-7), data was aggregated. When the sample of individuals included in the study was less than ten, verification was sought from Kate Granger's widower, Chris Pointon. Further research is warranted into the types of account holders posting to different social media and whether they are verified accounts or potentially bots.
2. Public/ private space (pg. 3)	The distinction between public (e.g. an internet site) and private space (e.g. the person's home) is increasingly blurred, mainly due to the introduction of new technology. On the internet, this blurring is extended to an extreme because much internet communication is conducted in both a private and public location	Existing literature identified the theoretical risks associated with the secondary use of data from social media as; the variable perceptions of 'public' and 'private' spaces, difficulties preserving the privacy of data subjects, consent to the use of social media data in research and

Section	BPS guideline / recommendation	Application to this PhD research project
	simultaneously. Moreover, a communication perceived as private at the time might become public at a much later date, when the archive is made publicly available.	experimental interventions that influence emotions, attitudes and behaviours of data subjects without their knowledge. Measures taken to deal with such risks were identified as: Only posts freely published on public social media were considered in scope. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten, i.e. Kate Granger's widower, Chris Pointon. Data was aggregated to mitigate the risk of identifying individuals and no personal information was included in any associated publications. The exception to this is Kate Granger's widower, where consent was granted.
4. Informed Consent (pg. 4)	It is both possible and indeed relatively usual for individuals to access web sites without reading instructions, explanations or terms and conditions. Participants may nominally give consent but without actually reading the relevant information. In this sense, consent is provided but it is not informed consent. In other (non-IMR) contexts consent to use data may be sought after the event. In IMR this arrangement is not reliable as the researcher can be caught unawares if the participant leaves unannounced before consent has been obtained and is subsequently untraceable. This does not mean that obtaining informed consent is impossible via the internet, but it does impose special constraints for which particular techniques can provide partial solutions.	Only posts freely published on public social media were considered in scope. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten, i.e. Kate Granger's widower, Chris Pointon.
4. Levels of control (pg. 5)	A key distinction between traditional research contexts and researching online is the degree of control (or lack of it) over the research environment. In IMR researchers do not have control over who the participants are, the environmental conditions they are responding under,	All empirical studies conducted as part of this PhD research project (Chapters 5-8) were observational in nature. The research team sought no control over the research environment.

Section	BPS guideline / recommendation	Application to this PhD research project
	variations in the research procedure due to different hardware and software configurations, what other activities or distractions are taking place simultaneously, participants' feelings and motivation and how they react to the research materials.	
5. Withdrawal (pg. 6)	In accordance with the general code of conduct for psychologists, online participants must have, and be aware of, rights to withdraw at any stage during a study. In the absence of physical contact between researcher and participant, extra precautions may be needed in IMR since it is not possible to gauge participants' attitudes and concerns at a distance. Difficulties need to be anticipated and withdrawal procedures made clear and simple. Contact e-mail address(es) should be provided on information sheets presented both prior to and post participation to allow participants to contact the research team. However, caution should be exercised about the exact details provided in order to preserve researchers' own privacy and safety.	Only posts freely published on public social media were considered in scope and were aggregated and anonymised to increase individual privacy. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than 10, i.e. Kate Granger's widower, Chris Pointon. The right to withdraw was available to those providing access or consent, however the ability to do so became increasingly problematic after the publication of the research studies.
6. Debriefing (pg. 6)	Debriefing online may take the form of a debriefing page displayed at the end of the study. Researchers may also consider the possibility of providing feedback via telephone, e-mail or other means. For practical reasons however, researchers might wish to restrict themselves to research where debriefing can be adequately carried out with a short piece of text on a debriefing page. Participants should be made aware of the level of debriefing (e.g. whether they will be told test scores) as part of the informed consent procedure.	No debriefing was provided except in the case of the longitudinal study (Chapter 8) where Kate Granger's widower, Chris Pointon, was provided with a copy of the manuscript prior to submission for publication.
7. Deception (pg. 7)	Many psychology experiments use forms of deception, accompanied by later debriefing. There must be a clear and convincing argument for the use of deception online, which is only condoned if the research question can be seen to justify it. This would typically be determined in part by appropriate institutional ethics committees. In	Only posts freely published on public social media were considered in scope. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten,

Section	BPS guideline / recommendation	Application to this PhD research project
	qualitative research the most likely instance of deception is in 'lurking' on chat rooms/discussion groups, etc. where researchers join a group as a member but use this to collect data. Other examples include publishing a website that misrepresents its aim. Strong justification is needed if the research involves deliberate misrepresentation by a researcher. In addition, there may be difficulties with debriefing which, in most instances, would be required. Researchers should therefore give careful consideration to whether or not the research is best conducted online.	i.e. Kate Granger's widower, Chris Pointon Ethics approval was sought and given by the University of Edinburgh's IRB. There was no deliberate misrepresentation by the researcher during the conduct of the empirical studies described in this thesis.
8. Manipulation (pg. 7)	Researchers should be mindful that some manipulations may produce effects not easily negated by debriefing. This could be images, video clips or descriptive text, swear words or psychological measures such as mood inductions or threats to self-esteem. Mood induction procedures administered via the Internet can produce emotional changes in participants and threats to self-esteem can be manipulated by perceived performance levels which can cause an emotional response. In such circumstances, debriefing may not be adequate and participants may withdraw without seeing it. For practical reasons, researchers may not be immediately available to discuss participants' experiences with them or intervene in other ways. Therefore, it is recommended that research materials that have a significant potential to cause distress, offence or have lasting effects on participants not be used in IMR. When material is presented that is somewhat sensitive, but not expected to create more extreme reactions than those normally encountered in the participants' everyday lives, it is to be recommended that researchers proceed with the same level of care and caution as in a face-to-face research study.	The development of an online intervention based on the data collected was not in scope for the empirical, observational studies conducted as part of this PhD research project.
9. Protection of participants	Researchers must be clear about steps they need to take to safeguard the anonymity of participants' contribution	Described earlier.

Section	BPS guideline / recommendation	Application to this PhD research project
and researchers (pg. 7-8)	to the project both online and in paper form. It is the researcher's responsibility to ensure that all information that would be needed by participants in order to have sufficient knowledge about the research is given in a clear, accessible and understandable form. ….. Researchers need to be aware of legislation concerning 'spamming' and act accordingly in using e-mail (or other forms of direct electronic communication) to make requests for participants. Researchers should be aware that steps to improve validity (e.g. using cookies, which are small pieces of data placed on the respondents' computer, checking IPs, etc.) can be constituted as an invasion of privacy. Participants may not be aware that they can take technical steps to protect their privacy, such as disabling 'cookies' in their browser. Researchers need to consider how to present themselves via the internet in order to sensitively maintain professional boundaries online. For example, many psychologists have a personal as well as research related websites. Care should be taken to maintain a boundary between research/professional use of the internet and personal life, for example by using separate e-mail addresses for professional and personal use on the corresponding website.	
10. Data protection (pg. 8)	Data protection legislation (Data Protection Act, 1998) has an impact generally on research conducted by psychologists, particularly around issues of storing personal data. Personal information can be processed during IMR, if participants consent to the processing of their data. This is a sufficient criterion to meet the requirements of the Data Protection Act (1998), according to the Information Commissioner. For research conducted on the internet additional issues need consideration. For example, researchers are required to lodge a notification of their research. These Society guidelines should not replace the responsibility of	Only posts freely published on public social media were considered in scope.

Section	BPS guideline / recommendation	Application to this PhD research project
	the researcher to ensure that they are following current Data Protection requirements, which are subject to change.	

Table 13. Application of the AoIR's 'Ethical Decision Making and Internet Research' guidelines

Section	AoIR's guideline / recommendation	Application to this PhD research project
	How is the context defined and conceptualised?	The online communities and individuals who post content on social media may expect a variable degree of privacy with regards to their profiles, interactions and status updates/ questions/threads. Status updates can theoretically be traceable to specific individuals. Only posts freely published on public social media were considered in scope Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten, i.e. Kate Granger's widower, Chris Pointon Data was aggregated to mitigate the risk of identifying individuals and no personal information will be included in any associated publications. The exception to this was Kate Granger's widower, where consent is granted.
	How is the context (venue/participants/data) being accessed?	Only posts freely published on public social media were considered in scope. The researcher was an observer of the interactions and had no direct contact with individuals except when consent was sought as described earlier.
	Who is involved in the study?	People posting content about the specific NCD on social media – there may be variation in expectations about privacy, use of data, outcome of study, traceability etc Researcher – health information posted onto social media that can be viewed by

		anyone with access to the internet should be considered public. Only posts freely published on public social media were considered in scope.
	What is the primary object of study?	Text data available in posts, threads and discussions on social media. Status updates can be traceable to specific individuals through account ids or the content of the post and therefore data was anonymised and aggregated as previously described. Health related data might be perceived as having a higher degree of privacy by authors compared to data associated with other topics. This was despite the data being shared on public social media.
	How are data being managed, stored, and represented?	During the conduct of the empirical studies (Chapters 5-8), the data files were stored on a local hard drive and Dropbox. Data files were shared amongst the research team via the university email system and disposed of by ‘shredding’ them electronically, following successful publication of the research study in a scientific journal. Data included in the publication of the research study, were aggregated and anonymised through the removal of reference to any account ids.
	How are texts/persons/data being studied?	Only posts freely published on public social media were considered in scope. Data extracted and analysed in the empirical studies were historical in nature. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten, i.e. Kate Granger’s widower, Chris Pointon.
	How are findings presented?	There is a risk that the participant perception of public and private space may change over time, which can be mitigated by the participant directly through their activity on social media.

		The right to withdraw was available to those providing access or consent, however the ability to do so became increasingly problematic after the publication of the research studies.
	What are the potential harms or risks associated with this study?	Only posts freely published on public social media were considered in scope, however individuals may experience negative exposure due to the content shared on public social media that is not directly associated with the studies and is outside the control of the research team. Rigorous application of ethical process and guidelines for conducting the empirical studies are described in the PhD thesis (Section 4.4). The risk of negative societal perception of online communities and research involving the secondary use of data from social media, may result in tighter controls, regulations and guidelines that inform the conduct of future research. Verification of identity is a known limitation of conducting research that involves data from social media. In order to mitigate against this risk and its potential impact on the validity of results from the empirical studies, data was aggregated. When the sample of individuals included in the study was less than ten, verification was sought from Kate Granger’s widower, Chris Pointon.
	What are potential benefits associated with this study?	Those that benefit from this research include: • The research community, through contributing to a growing body of knowledge that can inform future research and interventions • People affected by the condition, and those supporting them (e.g. healthcare teams, family, carers, clergy), should consider online communities as an additional source of social support during times of crisis. • Design of innovative palliative care services.

	How are we recognising the autonomy of others and acknowledging that they are of equal worth to ourselves and should be treated so?	Only posts freely published on public social media were considered in scope. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten, i.e. Kate Granger's widower, Chris Pointon.
	What particular issues might arise around the issue of minors or vulnerable persons?	All users of social media are assumed to be over the age of 16 given the terms and conditions for creating an account. Minors were not the intended scope of these empirical studies. Health-related data might be perceived as having a higher degree of privacy by authors, compared to data about other topics despite it being shared publicly on social media.

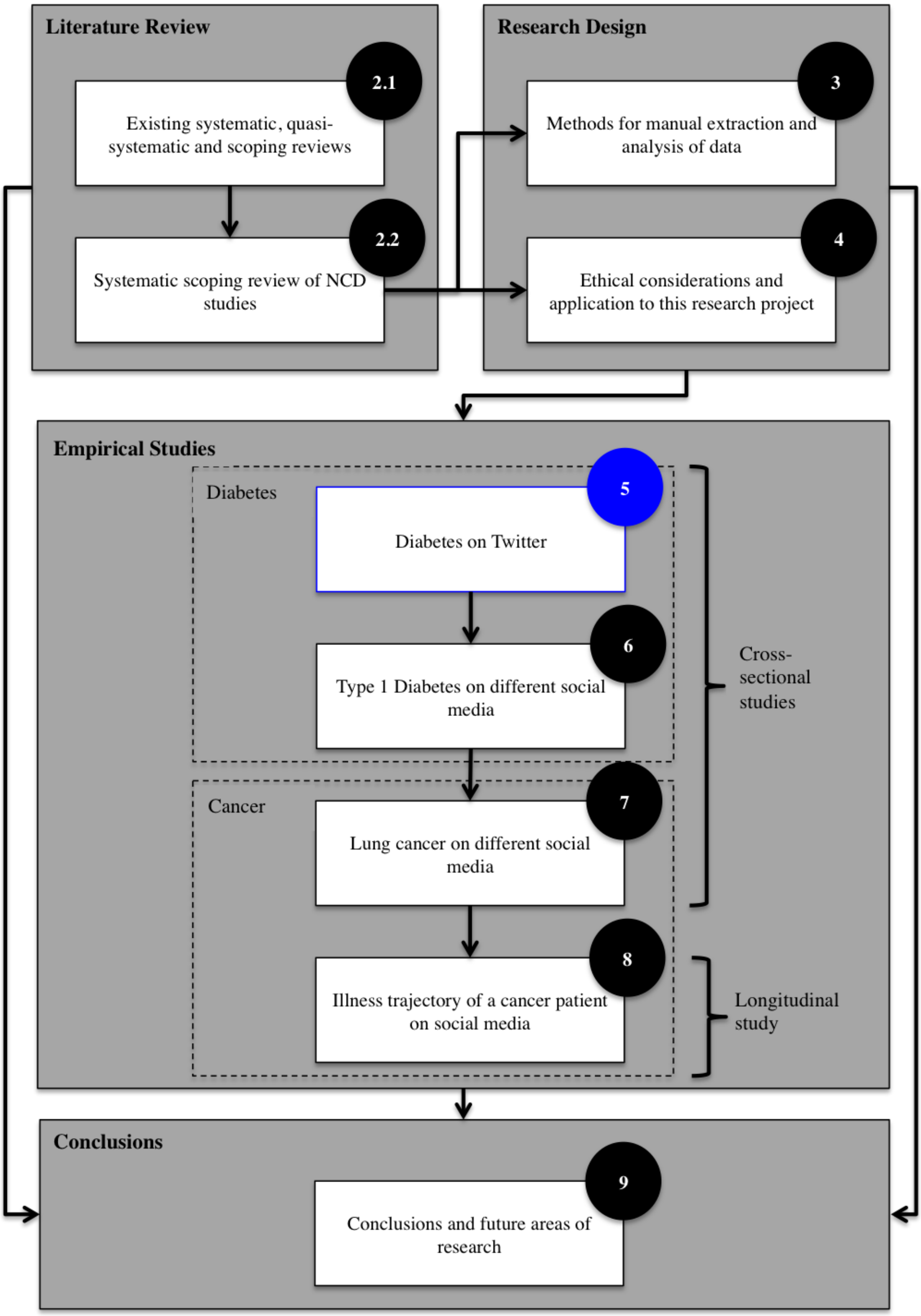
Table 14. Application of the NIHR's 'Guidance on the use of social media to actively involve people in research'

Section	NIHR's guideline / recommendation	Application to this PhD research project
8. Ethics (pg. 10)	Make sure that you are clear about whether you plan to use social media to actively involve people or to recruit them as participants or as 'subjects' of the research. Ethics committee approval is not needed if you are involving people in the design or conduct of research, but is needed if they will be participants of your research.	Social media was the intended source of data for the empirical studies (Chapters 5-8) No individual participants were actively recruited for these empirical studies. The self-selection of participants was due to the dependency on having a social media account in order to post a message. Consent was sought from group administrators for access to closed online communities (i.e. closed Lung Cancer Support Group on Facebook) or when the sample of individuals included in the study was less than ten, i.e. Kate Granger's widower, Chris Pointon.
	Make sure that you check any guidance produced by your institution about the use of social media. Some professional bodies and regulators also have guidance on this.	Guidance from the ESRC (Table 11), BPS (Table 12) and AoIR (Table 13) were reviewed and described in this chapter. Ethics approval was sought and given by the University of Edinburgh's IRB

		for the conduct of these studies.
	Familiarise yourself with some of the ethical issues relating to the use of social media.	Described in Chapter 4.

It was determined that the data available on social media, for the conduct of the empirical studies described in this thesis, exist in the public domain and therefore can be mined for research purposes without the informed consent of the authors [329]. Specific exceptions to this are when accessing a closed online community as described in Chapter 7 and in the extraction, analysis and publication of a single patient’s posts as described in Chapter 8. In order to protect the anonymity of content authors, their account id and any reference to other account IDs (e.g the use of @) were removed. A Level 1 Self Assessed Research Ethics Application was submitted and approved in 2014 for the empirical studies of diabetes detailed in this thesis (Chapters 5 and 6). This was later expanded to include empirical studies of cancer (Chapters 7 and 8). This adhered to the Research Ethics Policy of the University of Edinburgh’s School of Health in Social Sciences.

As was discussed in Chapter 2, the literature associated with the secondary use of data from social media for research purposes is evolving. Not surprisingly, so too are the ethics guidelines. In 2017, the BPS published an updated version of its guidance for researchers [306]. Version 3.0 of the AoIR’s guidelines followed this in 2019 [222]. Neither of these were taken into account during the conduct of the review described in Section 4.2 or in the empirical studies (Chapters 5-8), as they preceded the publication of the latest guidelines.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

5. To what extent are diabetes-related postings on Twitter relevant to the clinical condition and what topics and intentions are represented in these posts?

This chapter describes the first of four empirical studies that were conducted over a four-year period as part of this PhD research project. Conducted during the first year of this PhD research project (2013-2014), this pilot study enabled familiarity with the data posted on social media in relation to a specific NCD to be gained, as well as helping inform the research design. The results of this study were presented at the Stanford Medicine X conference in 2014 as a research presentation (Appendix A.1.1).

5.1 Context and aim of the pilot study

As highlighted in the results of the systematic scoping review of the literature (Section 2.2), Twitter has been used as a source of data in numerous public health research studies aimed at understanding how people seek and receive support in relation to conditions such as cancer [330-332], dementia [333] and epilepsy [334, 335] and how these technologies can be deployed to disseminate public health information about vaccination [336], sexual health [337], alcohol consumption [338, 339] and smoking [340, 341], for example. Twitter data has also been used extensively to monitor infectious disease outbreaks [342-346], including Ebola [347] and Zika [348].

Diabetes is a NCD and chronic condition, which occurs when the pancreas does not produce enough insulin (Type 1) or the body cannot effectively use the insulin it produces, resulting in an increased level of glucose in the blood (Type 2) [349]. Gestational diabetes is similar to Type 2 diabetes and is characterized by hyperglycaemia during pregnancy. The number of people diagnosed with diabetes has increased significantly over the past 20 years, from 151 million people worldwide in 2000 to 382 million in 2013 [350-354]. Projections indicate that it will be the seventh leading cause of death by 2030 [349].

Figures published by the Pew Research Center in 2013 indicate that 18% of online adults in the United States use Twitter compared to other social media such as Facebook (71%), Instagram (17%), Pinterest (21%) and LinkedIn (22%) [355] and that in 2010 diabetes was the health condition ninth most commonly-searched by people looking for information online [356]. As discussed in earlier chapters, Twitter is a micro-blogging platform where, in 2013, 500 million tweets were posted per day [357] by

218 million active users [358]. In 2013 tweet length was limited to 140 characters and was later expanded to 280 characters from 2017. It is a popular source of data for research studies and goes so far as to promote access to data for research purposes and innovation [226]. Twitter's limitations in the research context are widely reported. They include sampling method for accessing the data, the volume of data available as well as self-selecting nature and provenance of contributors. This can lead to challenges with bias and topic relevance, suggesting that Twitter can be a 'noisy' source.

This pilot study aimed to assess the relevance of diabetes-related postings on Twitter to the clinical condition and the topics and intentions represented in the posts. To guide the conduct of this pilot study, the following research questions were posed:

i. For what purpose do account holders use Twitter when referring to the clinical condition of diabetes?
ii. What is the relative occurrence of different types of diabetes within Twitter posts?
iii. Can statistical significance and relational meaning be determined from the content and intentions represented in the diabetes-relevant postings on Twitter?

5.2 Methods

The re-purposed PRISMA guidelines (Chapter 3) and ethical considerations (Section 4.2) described in earlier chapters were applied to this empirical study. Each phase of the PRISMA guidelines is detailed in subsequent sections of this chapter.

5.2.1 Identification

The stream of tweets was accessed using the standard, freely available 1% random sample that Twitter provides through their publicly accessible API stream [359]. From the Twitter API stream, a sample of tweets that were posted between 0:00:00 (UTC) on 1 September 2013 and 23:59:59 (UTC) on 30 September 2013 and included the term 'diabetes' was downloaded using a script prepared by Dr. Miles Osborne (co-supervisor). The timeframe for the sample of tweets was selected due to its convenience and to avoid undue bias or influence created by annual events such as World Diabetes Day (14th November). Spam was identified as tweets that did not include the term 'diabetes' but had been included in the sample or were generated by automated bots. Many of these referred to the purchase and availability of Viagra and were manually removed during the initial sampling. The data extracted included the date and time of posting in UTC, the Twitter account ID and the text in the tweet.

5.2.2 Screening

The sample of tweets was screened for relevance and inclusion based on the criteria in Box 5.

Box 5. Tweet inclusion and exclusion criteria for pilot study

Inclusion

Types of data: Data posted on the social media platform Twitter.
Language: English
Publication timeframe: Tweets posted between 0:00:00 (UTC) on 1 September 2013 and 23:59:59 (UTC) on 30 September 2013
Content of tweets: Tweets containing the word ‘diabetes’ and those that are classified as ‘health-related’ where ‘health-related’ is defined as:

- relevant to the prevention, diagnosis and treatment of the condition
- relevant to the human patients, professionals, organisations and general public impacted by the condition
- includes activities such as support seeking, information dissemination and opinion dissemination
- reference to fundraising activities and charitable causes associated with diabetes charities

Exclusion

Language: Not English
Content of tweets: Tweets not containing the word ‘diabetes’ and those that are classified as non-health-related as they:

- are intended as disparaging jokes or humour that attempt to denigrate people living with diabetes
- use diabetes as a synonym for sweet
- use diabetes in a context that is unrelated or where the relevance to health cannot be inferred

Any spam or re-posts that were missed during the initial sampling were identified and manually removed during the screening stage. Re-posts are defined as tweets that contain identical content and are posted by the same account ID at exactly the same time. Re-tweets are tweets which contain identical content but have been posted by a different account id. Re-tweets were classified as legitimate data points, as they reflect the decision of an independent user to communicate information and were, therefore, not removed from the sample. URL links included in the tweets were not reviewed during the screening.

5.2.3 Qualitative analysis

Tweets that met the inclusion criteria were then subjected to a qualitative content analysis akin to methods used in systematic reviews of qualitative research [360], in order to determine their form, the type of information contained within them, the implied intention behind the tweet and the user group for which the tweet was relevant (Figure 29). These four dimensions of the qualitative analysis are defined in Table 15.

Figure 29. Process steps and outcome of the analysis from pilot study

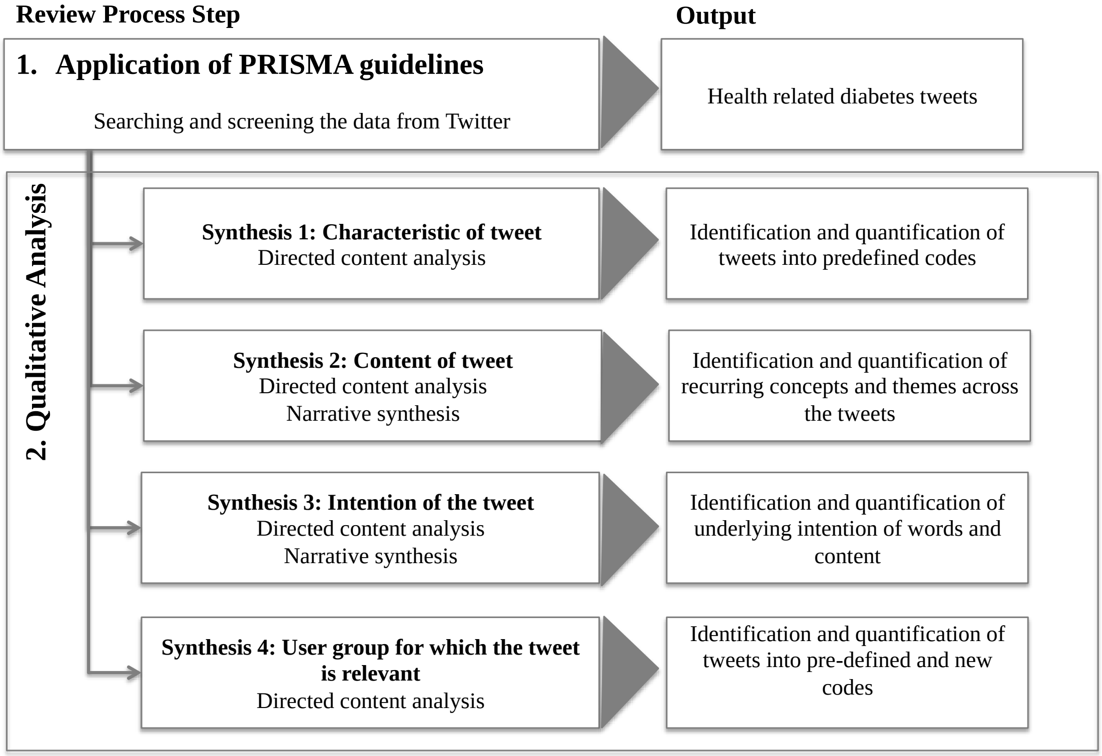


Table 15. Definition of dimensions for pilot study

Dimension	**Definition of dimension**
Characteristics of the tweet	These describe the functional characteristics and attributes of the tweet and have no pre-defined relation to the content of the post or the intention behind the post.
Content of the tweet	These describe the themes and topics raised within the content of the post (i.e. the 'what').
Intention of the tweet	These describe the purpose of the tweet and the contribution of the account id to discussion amongst the Twitter community (i.e. the 'why').
User group for which the tweet is relevant	These describe the group of people or organisations for which the tweet refers to or is relevant (i.e. the 'who').

Posts were initially categorised according to whether they represented an original tweet, a duplicate (i.e. they contained the same content as another tweet, but were posted at a different time or by a different account ID) or a re-tweet. Tweets authored by an organisation or group of individuals were then differentiated as being separate from those originating from a sole individual. The origin of tweets was manually categorised by viewing the profile page of the Twitter account ID. To the best of the researcher's knowledge, those tweets that were generated by automated bot accounts were removed as spam during the extraction process and were not included in the final sample. Given the binary categorisation of the tweets, those tweets that were not authored by an organisation or group of individuals were defined as authored by an individual.

The narrative content of the tweets was then analysed according to their fit with a pre-defined set of categories. These included diet, exercise, symptoms, risks, causes, diagnosis, complications, cure, products, procedures and self-help books, as well as research studies and news media, as described in previously published studies [75, 143]. These categories encompassed different levels of granularity and did not distinguish between the different forms of diabetes. Any text that could not be categorised into these original categories was given a new classification. The classification framework was iterated such that initial classifications were re-grouped and collapsed. Narrative synthesis using these categories was then used to interpret the results, the outcome of which was the identification and quantification of recurring concepts and descriptive themes across the tweets.

The intentions of the tweets were content analysed based on the Interaction Process categories specified by Bales [361] and Roter [362] and the functional roles of group members specified by Benne and Sheats [363]. Interaction Process Analysis is used to classify an individual's contributions to group discussions according to their form and function. The relevance of the tweet to a specific user group was also classified according to pre-determined criteria gleaned from previous studies, which isolated the user group through their sampling criteria. New classifications were again introduced if the tweet did not fit the initial coding framework. Classification codes were not mutually exclusive and a single tweet could be classified using multiple classification codes.

5.2.4 Quantitative analysis

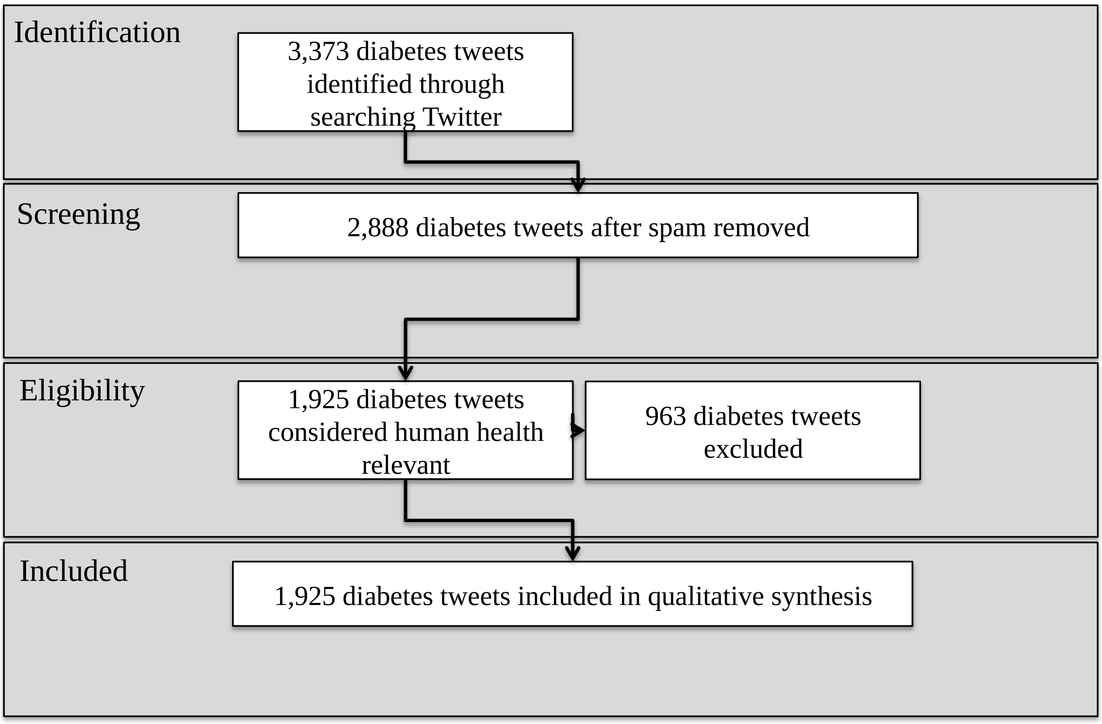
The outputs of the screening and content analysis were tabulated and analysed using SPSS v19 software. This provided descriptive statistics pertinent to the research questions for this pilot study and a hierarchical cluster analysis was conducted to determine the correlations between the content and intentions represented in the tweets. Hierarchical cluster analysis is a method of cluster analysis that groups similar data objects into groups called clusters. The endpoint is a set of clusters, where each is distinct from other clusters, and the data objects within each cluster are statistically similar based on the Euclidean distance and linkage criteria. Due to the dichotomous nature of the tweet classifications and the reduced likelihood of chaining, which can occur when single linkage clustering methods are applied, the complete linkage furthest neighbor method and phi-4 correlation where chosen for the hierarchical cluster analysis [364].

5.3 Results

A sample of 3,373 tweets was downloaded from the Twitter API and was identified as including the term ‘diabetes’. A sub-sample of 2,888 tweets remained following the removal of any non-English posts and spam. Of these, 1,925 were then identified as being health-related, based on the predefined inclusion criteria; those 1,925 tweets were then used as the basis for the qualitative and quantitative analyses.

Identification was made of 3,373 tweets that included the word ’diabetes’. Those that were not in English or identified as spam or a re-post were excluded before the sample of diabetes tweets was screened for relevance and the pre-defined inclusion and exclusion criteria applied. The number of eligible tweets included at each stage is shown in Figure 30. Examples of tweets that were excluded are: *“been playing #candycrush so long i think my iphone has developed diabetes!”* and “*girl you soo fine. eye candy; my eyes got diabetes”.*

Figure 30. Number of eligible tweets at each stage of the pilot study



5.3.2 Results of qualitative analysis

The pre-defined and enhanced classification framework structure produced from the mixed-method analysis is illustrated in Figure 31. Definitions for each of the 24 classification codes are provided in

Table 16 and are grouped into four dimensions; characteristics of the tweet, content of the tweet, user group for which the tweet is relevant and intention of the tweet as previously described.

Figure 31. Classification framework applied during the pilot study

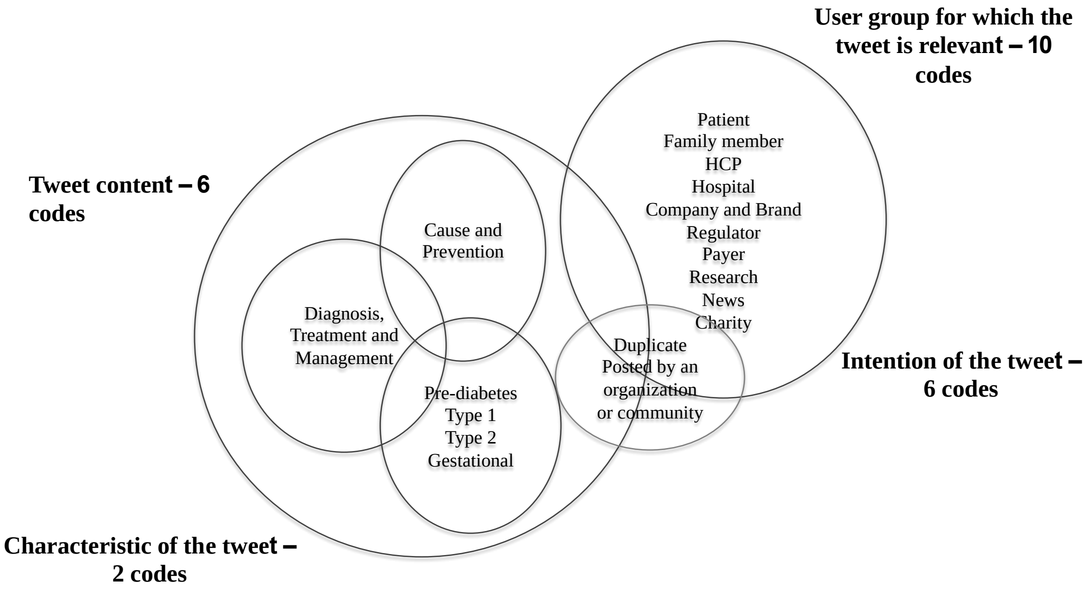


Table 16. Classification code framework applied during the pilot study

	Classification code	Code definition
Characteristics of the tweet		
1	Re-tweet or duplicate	A tweet that is a re-tweet of content already posted on Twitter or is the same content posted by a different account ID with a different timestamp.
2	Posted by an organisation	A tweet posted by an organisation or community. Note: Due to the binary nature of the classification, those tweets that were not classified as being posted by an organisation were defined as being posted by individuals.
Content of the tweet		
3	Diagnosis, treatment and management	A tweet that refers to the diagnosis, treatment or management of the condition once an individual has been classified as a patient for this condition. Examples of this include reference to treatment programmes, medication, links to other conditions, complications, medical devices, alternative therapies and the patient's electronic health record.
4	Cause and prevention	A tweet that refers to the cause or a means of preventing the condition for those who have not yet been diagnosed.
5	Pre-diabetes	A tweet that refers to the concept of pre-diabetes.
6	Type1	A tweet that refers to Type 1 diabetes.
7	Type 2	A tweet that refers to Type 2 diabetes.
8	Gestational	A tweet that refers to gestational diabetes.
Intention of the tweet		
9	Support seeking	A tweet that expresses the need for emotional or financial support.
10	Information giving	A tweet that provides orientation, repeats, clarifies, confirms.
11	Information seeking	A tweet that asks for orientation, repetition, clarification, confirmation.
12	Opinion giving	A tweet that provides evaluation, analysis, expresses feeling, wish.
13	Opinion seeking	A tweet that seeks evaluation, analysis, expression of feeling.

14	Ridicule	A tweet that ridicules or is intended to insult either directly or through humour.
User group for which the tweet is relevant		
15	Patient	A tweet that is relevant to or contains reference to diabetes patients.
16	Family member	A tweet that is relevant to or contains reference to a family member of a diabetes patient.
17	HCP	A tweet that is relevant to or contains reference to health care professionals.
18	Hospital	A tweet that is relevant to or contains reference to a hospital or medical centre.
19	Company or brand	A tweet that is relevant to or contains reference to a commercial organisation or brand such as a life sciences or consumer products company.
20	Payer	A tweet that is relevant to or contains reference to payers or medical insurance providers.
21	Regulator	A tweet that is relevant to or contains reference to health care and pharmaceutical industry regulators.
22	News	A tweet that contains reference to a news story.
23	Research	A tweet that refers to an ongoing or published research study and is relevant to a medical or public health research institute.
24	Charity	A tweet that is relevant to or contains reference to a diabetes charity organisation and its associated activities.

5.3.3 Results of quantitative analysis

The descriptive statistics (Table 17) describe the frequency of each of the 24 classification codes relevant to the sample of 1,925 diabetes tweets. The results highlight that 43.9% (n= 845) of the tweets were related to the diagnosis, treatment and management of the condition, compared to 26.2% (n=518) related to cause and prevention, and that 16.5% (n=318) were relevant to patients. Type 2 diabetes was more commonly discussed (12.1%, n=233) compared to Type 1 diabetes (6.0%, n=116) and gestational diabetes (1.6%, n=31) and the majority of tweets, 48.5% (n=934), were classified as information giving in nature.

Table 17. The frequency of each of the classification codes for the sample of diabetes tweets included in the pilot study

	Classification Code	**Number of health-related tweets (% of sample size:1,925)**	**Examples of the different categories of tweets**
Characteristics of the tweet			
1	Re-tweet or Duplicate	397 (20.6%)	
2	Posted by an organisation	397 (20.6%)	
Content of the tweet			
3	Diagnosis, treatment and management	845 (43.9%)	‘my #bgnow is 130. this makes me very, very happy. #diabetes’ ‘still at hospital... supporting our friend who's sick and on dialysis due to diabetes’
4	Cause and prevention	504 (26.2%)	‘healthy lifestyle success: "what causes diabetes and treatment options"’
5	Pre-diabetes	29 (1.5%)	‘higher "doses” of #exercise are better in #prediabetes’
6	Type1	116 (6.0%)	‘r.i.p. pancreas.. 3years today #diabetic #type1diabetes #diaversary’ ‘#type1diabetes. when my pump doesn't go well with my outfit...’
7	Type 2	233 (12.1%)	‘type 2 diabetes drugs in the pipeline’ ‘antidepressants may increase risk of type 2 diabetes’
8	Gestational	31 (1.6%)	‘anyone had gestational diabetes in #pregnancy? would like to hear more from someone who had it please! #mumtobe’
Intention of the tweet			
9	Support-seeking	223 (11.6%)	‘please sponsor me and rach to jump out of a plane! we're major wimps and may need to be pushed surely that's worth a quid!! @diabetesuk’
10	Information-giving	934 (48.5%)	‘a new class of diabetes drug doesn't increase heart-attack risk, two large studies show’
11	Information-seeking	48 (2.5%)	‘does colostrum help prevent type 2 diabetes?’
12	Opinion-giving	693 (36.0%)	‘i tried the abc diet but my diabetes stopped me from fasting. i can't eat below 500’ ‘check out this amazing blog by a young girl with diabetes, rising above and staying a fashionable tween.’
13	Opinion-seeking	69 (3.6%)	‘what do you feel is a realistic goal to be in range (75-140mg/dl) with your sensor?? #diabetes’
14	Ridicule	42 (2.2%)	‘i'm so glad i got this diabetes cookbook, i didn't know what to do with all the extra diabetes i had laying around in the food pantry.’ ‘it's not that diabetes, heart disease or obesity runs in your family. it's that no one runs in your family’

User group for which the tweet is relevant			
15	Patient	318 (16.5%)	‘big breakfast may be best for diabetes patients’ ‘my diabetes supplies cost enough. don't need to pay added tax for my insulin pump too’
16	Family member	64 (3.3%)	‘my mom has diabetes and i just found out ,i feel like a bad daughter’ ‘today would of been my uncles birthday, he passed away last year in his 50's. an infection in his leg led to an amputation #diabetes’
17	HCP	81 (4.2%)	‘diabetes nurse educator - baylor health care system: (#grapevine , tx) #healthcare #job #jobs #tweetmyjobs’ ‘india has too few cardiac, diabetes specialists’
18	Hospital	23 (1.2%)	‘man raises hundreds for the diabetes unit which 'saved his life'’ ‘hospital failed to monitor diabetic woman’s blood sugar’
19	Company or brand	79 (4.1%)	‘nejm publishes takeda examine cardiovascular safety outcomes trial in type 2 diabetes patients’ ‘sanofi says to pull u.s. diabetes drug application’
20	Payer	27 (1.4%)	‘diabetes may consume more than half of china’s annual health budget’ ‘health fact: the estimated cost of treating diabetes and its complications in the world is estimated as $215-375 billion.’
21	Regulator	27 (1.4%)	‘fda advisors recommend relaxing cv requirements for diabetes drugs’
22	News	129 (6.7%)	‘bloomberg news: 11.6% of chinese adults, or 114 million people, have diabetes’
23	Research	135 (7.0%)	‘increasing muscle mass by 10% will result in a 12% drop in risk of developing pre-diabetes -journal of clinical endocrinology’ ‘ ‘study in bmj suggests fruits, especially #blueberries cut risk of #type2 #diabetes up to 26%’
24	Charity	100 (5.2%)	‘i'm fundraising for diabetesuk on justgiving. please #sponsorme’

In order to understand whether a relational meaning exists between the classification codes, a hierarchical cluster analysis was performed to group tweets with similar characteristics, therefore aiding interpretation and enabling meaning behind the tweets to be inferred across multiple dimensions. By calculating the correlation between the different content, intentions and user group categories, this enabled the strength of the relationship, i.e. the proximity matrix, between each of these categories to be determined (Table 18). The results of the hierarchical cluster analysis are illustrated in the dendrogram

(Figure 32). The strength of relationship between the 24 classification codes ranges in value from -0.511 (CausePrevent and DiagTreatMgt) to 0.383 (SupportSeeking and Charity). The shorter the Euclidean distance the weaker the linkage. SupportSeeking and Charity were identified as having the closest proximity with 0.383 suggesting a linkage between these two categories of tweets, which is graphically illustrated in the dendrogram. CompanyBrand and Regulator (0.267), Research and InfoGiving (0.184), CausePrevent and Type2 (0.156), HCP and Hospital (0.121) and Patient and Type 1 (0.105) were also identified as being linked.

Figure 32. Dendrogram from cluster analysis of categories represented in the pilot study

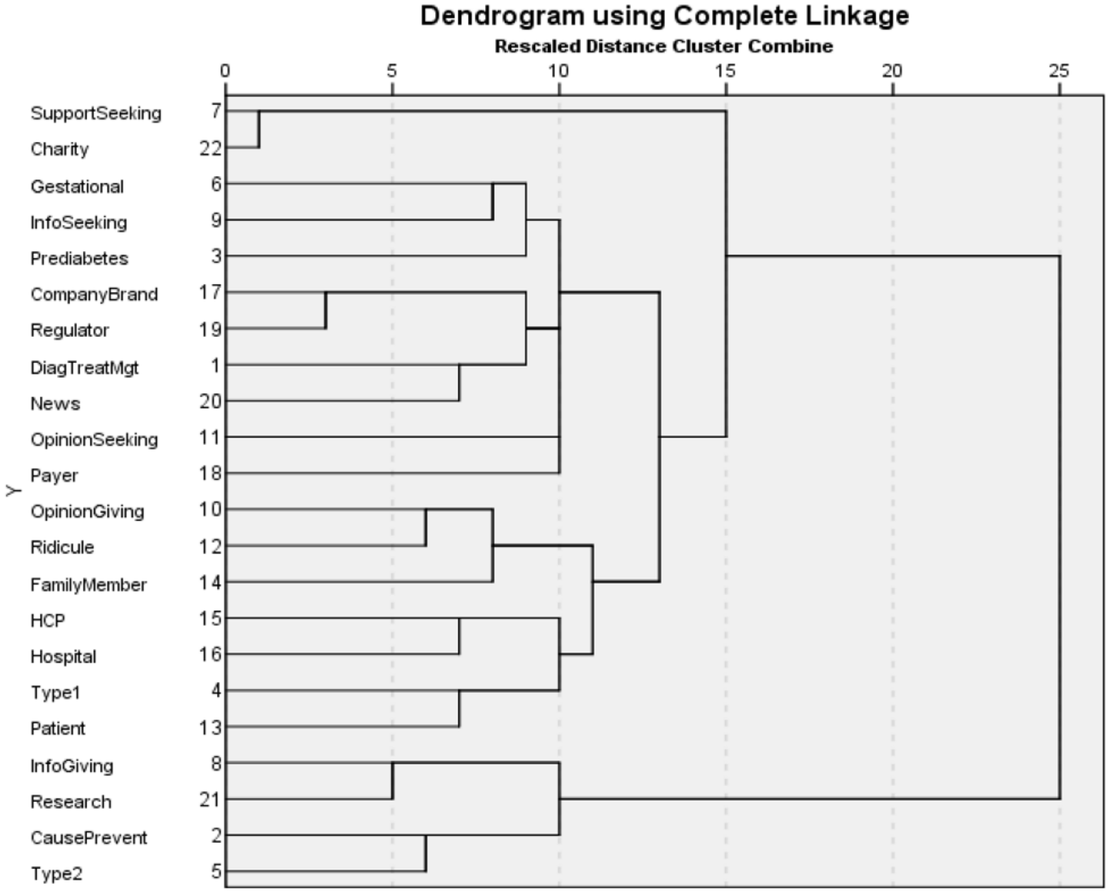


Table 18. Proximity matrix of categories included in the pilot study

Proximity Matrix

Case	Matrix File Input																					
	DiagTreatMgt	CausePrevent	Prediabetes	Type1	Type2	Gestational	SupportSeeking	InfoGiving	InfoSeeking	OpinionGiving	OpinionSeeking	Ridicule	Patient	FamilyMember	HCP	Hospital	CompanyBrand	Payer	Regulator	News	Research	Charity
DiagTreatMgt	1.000	-.511	.024	.031	-.103	-.018	-.056	.052	.023	-.021	.004	-.068	.041	-.028	.031	.018	.057	.005	.063	.093	.005	-.165
CausePrevent	-.511	1.000	-.003	-.121	.156	-.046	-.175	.048	-.036	.052	-.007	.024	-.188	-.030	-.118	-.066	-.039	-.039	-.051	-.056	.004	-.129
Prediabetes	.024	-.003	1.000	-.031	.035	.020	-.030	.056	.035	-.037	-.023	-.018	-.007	-.022	-.004	-.013	-.025	-.014	-.014	.002	.018	-.028
Type1	.031	-.121	-.031	1.000	-.007	-.032	.011	.038	.001	-.040	-.002	-.008	.105	.027	-.009	.012	-.052	-.011	-.012	.046	.085	.000
Type2	-.103	.156	.035	-.007	1.000	-.047	-.114	.161	.021	-.108	-.011	-.044	-.100	-.041	-.045	-.026	-.019	-.016	-.031	-.023	.030	-.058
Gestational	-.018	-.046	.020	-.032	-.047	1.000	.007	-.013	.060	-.007	-.024	-.019	.001	.000	.016	-.014	-.005	-.015	-.015	-.017	-.034	-.029
SupportSeeking	-.056	-.175	-.030	.011	-.114	.007	1.000	-.348	-.058	-.234	-.052	-.043	.189	.098	-.018	.020	-.066	-.042	-.029	-.078	-.099	.383
InfoGiving	.052	.048	.056	.038	.161	-.013	-.348	1.000	-.150	-.667	-.148	-.131	-.134	-.126	.027	.037	.085	.031	.026	.177	.184	-.171
InfoSeeking	.023	-.036	.035	.001	.021	.060	-.058	-.150	1.000	-.121	-.031	-.024	-.036	-.011	-.034	-.018	-.016	-.019	.009	-.030	-.044	-.008
OpinionGiving	-.021	.052	-.037	-.040	-.108	-.007	-.234	-.667	-.121	1.000	-.145	.162	.020	.075	.001	-.033	-.039	-.003	-.025	-.110	-.107	-.107
OpinionSeeking	.004	-.007	-.023	-.002	-.011	-.024	-.052	-.148	-.031	-.145	1.000	.029	.020	-.020	-.012	-.021	-.011	.002	.025	.027	-.042	.056
Ridicule	-.068	.024	-.018	-.008	-.044	-.019	-.043	-.131	-.024	.162	.029	1.000	.020	.052	-.031	-.016	-.031	-.017	-.018	-.040	-.041	-.035
Patient	.041	-.188	-.007	.105	-.100	.001	.189	-.134	-.036	.020	.020	.020	1.000	.021	.076	.119	-.070	.021	-.017	-.018	-.050	-.066
FamilyMember	-.028	-.030	-.022	.027	-.041	.000	.098	-.126	-.011	.075	-.020	.052	.021	1.000	-.024	.034	-.038	.004	-.022	-.014	-.016	-.017
HCP	.031	-.118	-.004	-.009	-.045	.016	-.018	.027	-.034	.001	-.012	-.031	.076	-.024	1.000	.121	.023	-.024	.019	.017	-.037	-.025
Hospital	.018	-.066	-.013	.012	-.026	-.014	.020	.037	-.018	-.033	-.021	-.016	.119	.034	.121	1.000	-.023	-.013	-.013	.047	-.011	.039
CompanyBrand	.057	-.039	-.025	-.052	-.019	-.005	-.066	.085	-.016	-.039	-.011	-.031	-.070	-.038	.023	-.023	1.000	-.001	.267	.040	.016	-.048
Payer	.005	-.039	-.014	-.011	-.016	-.015	-.042	.031	-.019	-.003	.002	-.017	.021	.004	-.024	-.013	-.001	1.000	-.014	.041	-.032	-.027
Regulator	.063	-.051	-.014	-.012	-.031	-.015	-.029	.026	.009	-.025	.025	-.018	-.017	-.022	.019	-.013	.267	-.014	1.000	.039	-.033	-.028
News	.093	-.056	.002	.046	-.023	-.017	-.078	.177	-.030	-.110	.027	-.040	-.018	-.014	.017	.047	.040	.041	.039	1.000	.033	-.053
Research	.005	.004	.018	.085	.030	-.034	-.099	.184	-.044	-.107	-.042	-.041	-.050	-.016	-.037	-.011	.016	-.032	-.033	.033	1.000	-.036
Charity	-.165	-.129	-.028	.000	-.058	-.029	.383	-.171	-.008	-.107	.056	-.035	-.066	-.017	-.025	.039	-.048	-.027	-.028	-.053	-.036	1.000

5.4 Discussion

The results from the systematic scoping review (Section 2.2) indicated that in 2013 a study had yet to be published that sourced data about diabetes from Twitter. The results from this pilot study indicate that 1925 of the 2888 (66%) screened sample of English language tweets, which include the word ‘diabetes’ are relevant to the clinical condition. This suggests that Twitter, despite being a ‘noisy’ source of data, can be considered a valid source of insight for the purposes of health research into the condition. Explaining why it continued to be used as a source of data for research studies investigating diabetes that were published after the completion of this pilot [138, 148, 151].

What account holders use Twitter for when referring to the clinical condition of diabetes

The first research question associated with this pilot study set out to determine the affordances offered by Twitter to account holders when referring to the clinical condition of diabetes. Posts referring to diabetes are common on Twitter and, while it remains a ‘noisy’ source, over 66% of tweets mentioning the term appear to be relevant to the clinical condition. The remaining 34% of tweets that were not considered relevant were classified as either unrelated to the condition, disparaging jokes or used ’diabetes’ as synonym or slang for sweet. To date, the identification of slang continues to present challenges to researchers analysing data sourced from social media [365].

In the pilot study as highlighted in Table 17, Twitter was used primarily for ‘information-giving’ (n=934, 48.5%) and ‘opinion-giving’ (n=693, 36.0%), whilst only a small proportion demonstrated ‘opinion-seeking’ (n=69, 3.5%) or ‘information-seeking’ (n=48, 2.5%). This supports previous studies that highlight that Twitter is primarily being used as a means of sharing and promoting information and opinions with other users about a range of different health topics [114, 330, 332-334, 337, 338, 340, 366-372], rather than as a two-way channel of communication [75, 373].

Despite the largely one-way dissemination function of Twitter highlighted in this pilot study, 11.6% (n=223) of the posts were in fact aimed at seeking support and using Twitter as a means of connecting with other members of the online community living with the same condition. Although unable to determine the reciprocated nature of the tweets, this finding suggests alignment with the results of a systematic review into the uses and benefits of social media for health communication [75], which determined that social media is used as an effective means of providing peer, social and emotional support to users. Although not specific to Twitter, this is further supported by findings from a study of the information and communication between users of a diabetes community on Facebook [374] where medical and lifestyle information and personal experiences, opinions and advice are shared amongst the community as a means of providing informational and emotional support to members.

The types of diabetes discussed on Twitter

The second research question associated with this pilot study set out to understand the relative occurrence of references to different types of diabetes within Twitter postings. The majority of diabetes-related tweets in the sample did not specify diabetes type, a pattern mirrored in the systematic scoping review described in Section 2.2. Of those that did, 12.1% (n=233) referred to Type 2 diabetes, 6.0% (n=116) to Type 1 diabetes and 1.6% (n=31) to gestational diabetes. Type 2 diabetes is the most common form of diabetes diagnosed in (90%) [375, 376] and therefore one might expect it to appear more frequently in the sample of tweets. Although this is the case in the pilot study, the percentage results from this study do not represent the true ratio of Types 2 to 1 diabetes occurrence amongst diabetes patients. One explanation for this is the user demographic of Twitter, with 66.2% of users under the age of 35 years [377]. Type 2 diabetes however is more prevalent amongst older people, who are less active users of social media, and Twitter in particular [355]..

The patient journey, including diagnosis, treatment and management of diabetes, is represented in 43.9% (n=845) of the tweets, while only 26.2% (n=504) concern the cause and prevention of diabetes. Findings from earlier research on online cancer communities [378, 379] concluded that individuals are more likely to use social media when they face a new diagnosis or a condition that requires continuous management, as is the case with diabetes, due to the uncertainty of living with the condition, a potential lack of personal knowledge about treatment options and the need for peer and emotional support during times of potential crisis. In Hamm et al.’s study, self-care was also identified as the main topic for which social media are used by patients in general [69]. Research published after the completion of the pilot study and identified in the systematic scoping review (Section 2.2) concur with these findings about social media use amongst the diabetes community, identifying informational support associated with current and newly approved treatment options, personal experiences and opinions about available medical care, self care through diets and exercise, support from family, as well as emotional support from peers [106, 112, 143, 148, 149, 151, 153, 189, 203].

Relational meaning between content and intentions as represented in the diabetes-relevant tweets

The third research question associated with this pilot study was aimed at understanding the relational meaning of the diabetes relevant posts on Twitter and what can be gleaned from the content and intentions represented in the tweets. Tweets seeking support had proximity of 0.383 with those referencing charitable organisations. This aligns with the motivational, emotional and financial support requested by users as they connect with charitable organisations aiming to reduce the prevalence of the condition or enable patients to achieve a better quality of life whilst living with their existing condition [380-382]. Tweets indicating an information-seeking intention have a close proximity (0.060) to those that refer to gestational diabetes, supporting studies that found expectant and new mothers frequently access social media sites as a means of understanding their condition [383, 384].

Tweets that refer to a brand or commercial organisation had a close proximity (0.267) with those that made reference to regulators and regulation. If the tweet referred to a brand-named drug or medical device that had either been approved or rejected for use by diabetes patients, it often also referred to the company that manufactures the product.

Diabetes tweets that referred to HCPs also often referred to a hospital or medical centre (proximity value 0.121), which is unsurprising given that HCPs are typically associated with medical institutions. Findings from other studies conducted in the USA suggest that HCPs use social media to communicate, share knowledge and develop skills [89] and that medical institutions, along with other organisations, also use social media as a means of advertising and recruiting for open positions [385-387]. The use of social media to advertise and recruit for open positions was evident in the 1925 tweet sample from this pilot study.

Tweets aimed at ‘information giving’ had a proximity value of 0.184 with those that referenced published or ongoing research in the field of diabetes, suggested that Twitter was being used as a means of disseminating science and innovation (n=135, 7.0%). This supported research charting the use of social media by researchers and academic institutions to disseminate their work with peers and the general public [388-390] and also relates to the new field of altmetrics, where web-based metrics, such as the number of tweets and re-tweets, are used as means of measuring the societal impact of scholarly material and research outputs beyond the traditional citation measure [391]. .

Tweets that referred to the causes and prevention of diabetes were closely linked (0.156) with those that made reference to Type 2 diabetes. This finding may be partly explained by the anxiety many people face when receiving a diagnosis of Type 2 diabetes, which is often associated with lifestyle factors (e.g. obesity, lack of exercise) and requires behavioural changes which can be confusing for patients [392]. This was further investigated in Holmberg et al.’s study into the content and conversational connections of diabetes and diets, which found that members of the online diabetes community advocate and share information about the condition and its associated risk factors as well as the interventions for reducing these risks [151]. In contrast, Type 1 diabetes may be perceived to be a ‘medical’ condition for which drug treatment is effective. Jones, Sinclair and Barnard’s 2013 systematic review found that young people with Type 1 diabetes were actively seeking advice on how to minimise the risks associated with alcohol consumption [49]. However, its findings were limited by only including a small sample of six studies that were identified. Despite the increased number of studies published since 2014, a gap still exists in the literature and further research is needed to achieve a more granular understanding of the relationship between NCDs such as diabetes and the associated risk factors, within the social media content.

Limitations

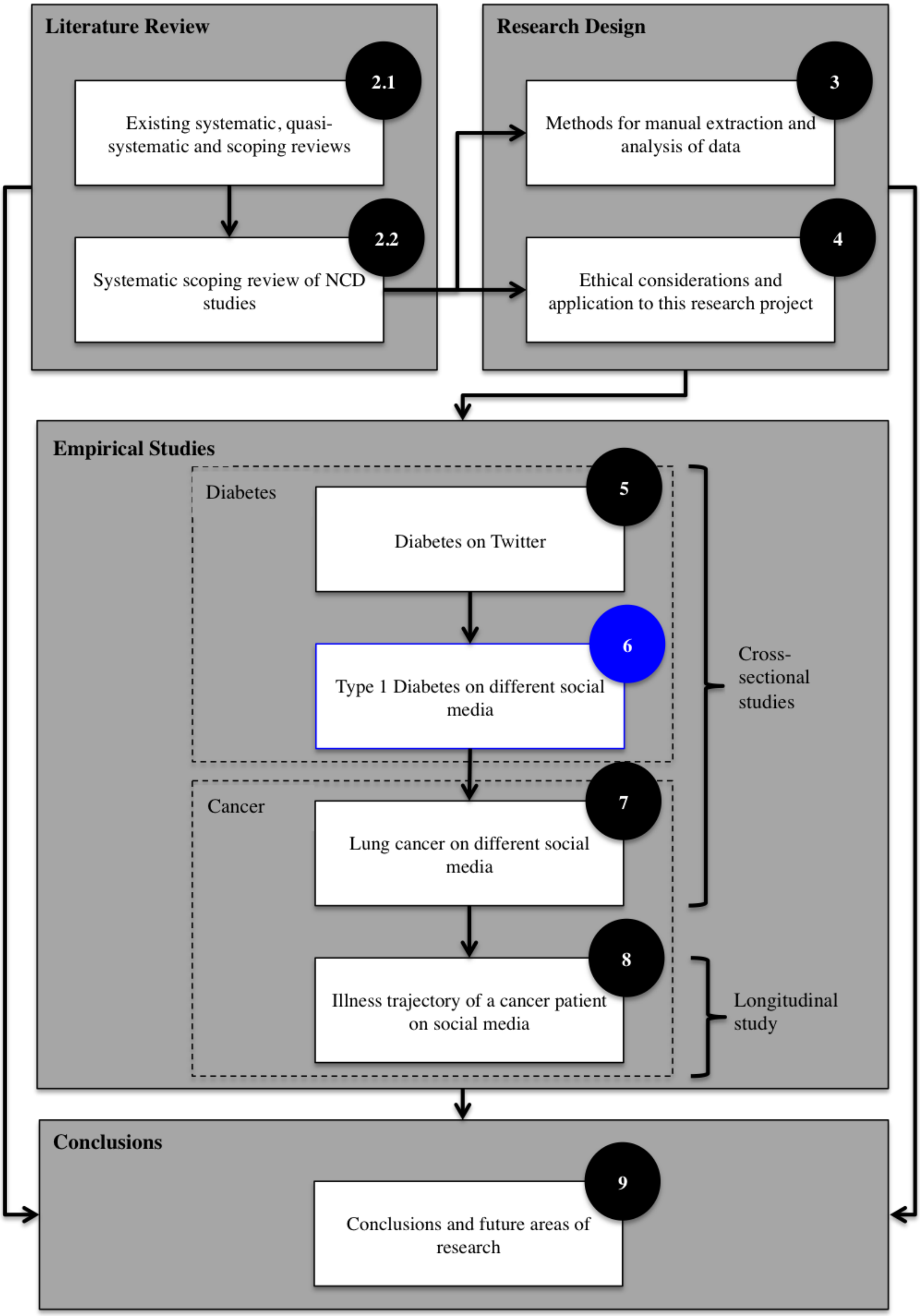
Limitations associated with this pilot study include those widely reported in other studies that source data from social media. They include the lack of publicly available information and clarity about how Twitter determines and ensures that the information which is provided through their API stream is a uniform and random, 1% sample of the Twitter Firehose, as well as whether the sample made publicly available can be considered representative of every single published tweet [393]. These unknowns could limit the extent to which generalisations can be extrapolated and made about the results. Other limitations include the self-selection of contributors, as a Twitter account is required in order to post a message and the sample may not therefore perfectly represent the population of people living with diabetes, given the 'digital health divide' [394] and the higher prevalence of diabetes in older age groups [395]. The level of detail provided within each of the posts also required a degree of interpretation by the researcher conducting the study. This can create a risk of interpretation bias and therefore example tweets for each category are provided in Table 17. Analyses of Twitter data also suffers from the presence of misinformation and provenance [396], as not all account holders are verified and 15%-20% of Twitter accounts are identified as bots [397]. This potentially false data can influence the results and findings from studies investigating the patient experience and hence have implications for the conclusions drawn. In addition, the cross-sectional nature of this study did not take into account variations in Twitter activity that may have occurred over time, nor did the study include all disease-specific hashtags or key words such as 'diabesity', 'diabetic' or 'T1D' [398], which may have contributed to the pool of relevant posts.

5.5 Conclusions

Conclusions from this pilot study suggest that Twitter is a potentially valid source of information about the clinical condition of diabetes, although a sizable proportion (n=963, 34%) of tweets containing the word diabetes were either irrelevant or misleading. Most (n=1627, 84.5%) diabetes-relevant tweets involved the giving of information and opinion by participants rather than seeking to derive, information or support (n=117, 6.1%). Diabetes-related commercial and charitable organisations are responsible for posting 1 in 5 of these messages (n=397, 20.6%). As such, Twitter may not necessarily be considered the optimal source of data for gaining insights into the patient experience of diabetes.

5.6 Implications for this PhD research project

Since this pilot study was conducted and presented at the Stanford Medicine X conference in 2014, 11 further studies that investigated diabetes on social media have been identified in the published literature [100, 106, 112, 121, 138, 148, 150, 153, 167, 189, 203]. Only one of these focused on Type 1 diabetes [138] whilst the remaining ten studies investigated Type 2 diabetes or diabetes more broadly. The objective of these empirical studies included: understanding the information needs of users [203] and how information is shared [106, 112, 121, 138, 148, 150], as well as adverse events [167] and treatments [100]. The data analysed in these studies were sourced from Facebook [100, 106, 189], Twitter [121, 138, 148] and WebMD [153]. Despite the fact that two of the identified studies sourced data from multiple discussion forums [112, 149] none went so far as to draw comparisons between different social media. Further research was therefore needed to understand how people affected by diabetes and other conditions, use different social media and what value these represent for public health research. The insights and conclusions from the pilot study were used to inform the empirical studies described in Chapters 6 and 7 of this thesis, and in doing so contributed to the body of knowledge associated with NCDs and their narrative on social media.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

6. Do people affected by Type 1 diabetes use different social media for different purposes and what are the implications for health researchers?

As highlighted in the results of the systematic scoping review (Section 2.2), Type 1 diabetes is under represented in the existing literature. Following the completion of the pilot study described in Chapter 5, it was identified that further empirical research was needed to understand the differences in utility and affordances offered by different social media platforms to people affected by Type 1 diabetes. Addressing this gap in existing literature and RQ5 is the focus of the second empirical study, conducted in 2014-2015, as part of this PhD research project and described in this chapter, the outcome of which was published as a short paper in the conference proceedings of the 2015 International Conference on Global Health Challenges in France (Appendix A.1.3).

6.1 Context and aim of the study

Type 1 diabetes is a chronic autoimmune condition, which occurs when the pancreas stops producing insulin. This results in increased levels of glucose in the blood, putting patients at long-term risk of heart disease, stroke, nerve damage, kidney disease and blindness. There is currently no known cure for Type 1 diabetes and those diagnosed are treated either by insulin injections, insulin pump therapy, islet cell transplantation or a pancreas transplant [399].

Previous research discussed in Section 2.2, as well as the pilot study described in Chapter 5, focused on single platforms such as Twitter, Facebook or condition specific online communities. The study described in this chapter went beyond such published research with a view to gaining insights into the social interactions that occur across and within social media. In 2015, there was no known published study to have compared Twitter, a Type 1 diabetes group on Facebook [400] and the Type 1 diabetes discussion forum on Diabetes.co.uk [401]. Motivated by this, the following research questions were considered, the answers to which will help patients better understand how, by using them, they can manage their conditions:

i. Do Type 1 diabetes patients use different social media platforms for different purposes?
ii. Which social media are successful at encouraging social interaction and support for patients?

The three social media embody different styles of social interaction. Twitter and Facebook are well-known, general-purpose platforms. Diabetes.co.uk is a condition specific discussion forum where users can create content and others can comment.

6.2 Methods

For the comparison of social media use by diabetes patients, the re-purposed PRISMA guidelines (Chapter 3) and ethical considerations (Chapter 4) described in earlier chapters were applied to this empirical study.

6.2.1 Identification and sampling

Twitter. A 1% sample of all available tweets was extracted from Twitter on the 3 October 2014. The Tweets were posted between 0:00:00 (UTC) on 1 September 2014 and 23:59:59 (UTC) on 10 September 2014 and were extracted by crawling the data through the standard publicly-available Twitter API using the pre-defined search terms diabetes, type 1 diabetes, t1 diabetes, t1d and type 1. The data extracted included the date and time of posting in UTC, the Twitter account ID and the text in the tweet. Re-tweets were identified and any duplicates and spam were removed.

Facebook. Using the Facebook search functionality, we searched for Type 1 diabetes and, in doing so, identified the largest Type 1 Diabetes Facebook group available. Known as the ‘Type 1 Diabetes Community’ this group was established in 2011 and is intended to be an open forum for people who have Type 1 diabetes to talk about anything they want. As of 4 October 2014, it had 36,671 likes and, on this same date, all wall posts and replies that were posted between 1 and 10 September 2014 were identified by viewing the storyline of historical posts. These posts, along with the author and the date of posting, were manually extracted for further analysis. Given that this is an open group, and anyone with a Facebook account can join, consent was not sought from the group administrators or members.

Diabetes.co.uk is a global diabetes community with over 125,000 members spanning all forms of the condition. The Type 1 discussion forum on Diabetes.co.uk was identified through the forum homepage and the list of discussion threads was then filtered based on the start date of 1 September 2014. All original posts and replies posted between 1 and 10 September 2014 were identified and manually extracted. Posts are accessible to all without need for a Diabetes.co.uk account.

6.2.2 Screening

Insights gained from the pilot study (Chapter 5) suggested that Twitter was a ‘noisy’ source of data. The sample of 1,433 English language tweets were therefore manually screened for relevance. Those that referred to Type 1 diabetes produced a sample of 66 posts, which were subsequently included in the categorisation stage. URL links included in the tweets were not reviewed during the screening.

Given the condition-specific focus of the Facebook group and the Diabetes.co.uk discussion forum, the total sample of 479 posts and 713 posts, respectively, were all considered relevant to Type 1 diabetes and therefore included in the categorisation stage.

6.2.3 Application of Bales’ IPA categories

The Bales IPA system [361] was first introduced in 1951 and has been widely used in public health research as a means of identifying and recording the nature, not the content, of group interactions. Bales identified and defined 12 categories of group interaction (Figure 33), each of which were considered when reviewing the Type 1 diabetes posts. Each post was considered to be a single unit of interaction and the categories were not considered to be mutually exclusive when applied to the sample of posts.

Figure 33. Definition of Bales IPA categories

Positive reactions	1	*Seems friendly,* shows solidarity, raises other's status, gives help, reward
	2	*Shows tension release,* jokes, laughs, shows satisfaction
	3	*Agrees,* shows passive acceptance, understands, concurs, complies
Attempted answers	4	*Gives suggestion,* direction, implying autonomy for others
	5	*Gives opinion,* evaluation, analysis, expresses feeling, wish
	6	*Gives information,* orientation, repeats, clarifies, confirms
Questions	7	*Asks for information,* orientation, repetition, confirmation
	8	*Asks for opinion,* evaluation, analysis expression of feeling
	9	*Asks for suggestion,* direction, possible ways of action
Negative reactions	10	*Disagrees,* shows passive rejection, formality, withholds help
	11	*Shows tension,* asks for help, withdraws out of field
	12	*Seems unfriendly,* shows antagonism, deflates other's status, defends or asserts self

6.3 Results

The results of the extraction and screening of Type 1 diabetes posts are summarised in Table 19, where the number of posts extracted from each of the three forms of social media, the number of original posts and replies included in the sample and their respective number of authors can be seen.

Table 19. Summary of Type 1 diabetes posts from 1-10 September 2014

Social Media	**Sample size**	**Screened sample size**				**Posts included in IPA**
		N (%) original posts	**N authors**	**N (%) replies**	**N reply authors**	
Twitter	1433 (based on 1% sample from Twitter Firehose)	66 (4.6%)	62	0 (0%)	0	66
Type 1 diabetes community on Facebook	479	16 (3.3%)	1	463 (96.6%)	310	479
Type 1 diabetes discussion forum on Diabetes.co.uk	713	41 (5.7%)	37	672 (94.2%)	123	713

The results of Bales IPA reveal differences in the nature of interactions between users of these three forms of social media. These are described in Table 20, where the percentage of posts relevant to the Bales IPA categories can be seen for each of the three forms of social media. It was also revealed that a single post within the Facebook community generated a higher response rate relative to others accounting for one fifth of the Facebook sample. It was therefore highlighted as a separate line item in Table 20, so as to avoid a potential skew of results within the Facebook dataset. The post '*Over/Under time again. 153. Are you over or under?*' was a request from the community moderator for members to post their current blood glucose levels. This post generated 101 responses, accounting for 21% of the total sample and therefore performing a role similar to that of an online survey.

Table 20. Application of Bales IPA to the Type 1 diabetes posts

Social Media	**Bales IPA Categories N (%)**					
	Shows tension	**Dramatises**	**Agrees**	**Gives Suggestions**	**Gives Opinion**	**Seems Unfriendly**
Twitter (based on 1% sample from Twitter Firehose)	5 (8%)	12 (18%)	0	13 (20%)	24 (36%)	0
Type 1 diabetes community on Facebook	19 (4%)	24 (5%)	43 (9%)	139 (29%)	158 (33%)	5 (1%)
Type 1 diabetes community on Facebook (minus the over/under post)	15 (4%)	23 (6%)	45 (12%)	140 (37%)	151 (40%)	4 (1%)
Type 1 diabetes discussion forum on Diabetes.co.uk	86 (12%)	50 (7%)	36 (5%)	128 (18%)	392 (55%)	21 (3%)
	Gives info	**Asks for info**	**Disagrees**	**Asks for Suggestions**	**Asks for Opinions**	**Seems friendly**
Twitter (based on 1% sample from Twitter Firehose)	19 (29%)	5 (8%)	0	3 (5%)	4 (6%)	12 (18%)
Type 1 diabetes community on Facebook	158 (33%)	29 (6%)	19 (4%)	5 (1%)	14 (3%)	177 (37%)
Type 1 diabetes community on Facebook (minus the over/under post)	57 (15%)	30 (8%)	19 (5%)	4 (1%)	11 (3%)	177 (47%)
Type 1 diabetes discussion forum on Diabetes.co.uk	228 (32%)	70 (10%)	7 (1%)	7 (1%)	43 (6%)	328 (46%)

6.4 Discussion

The results from this study suggest that there are indeed differences in the utility and affordances offered by the three different social media to people affected by diabetes. The 1% sample from Twitter had the highest absolute number of total posts at 1,433, supporting the findings from the pilot study (Chapter 5) that it is a ‘noisy’ source of data compared to other social media, as only 66 of the 1433 posts were relevant to the condition of Type 1 diabetes. Tweets that were excluded were either identified as spam, intended as disparaging jokes or humour that attempted to denigrate people living with Type 1 diabetes or used Type 1 diabetes in a context that was unrelated or where the relevance to health could not be inferred. Given that these results were based on a 1% sample from the Twitter firehose, the total number of tweets during the ten-day period can be extrapolated to 143,300.

The results also revealed that, despite being a popular social media, the Type 1 diabetes group on Facebook, with a total of 479 posts, was not the most actively used platform for members to discuss the condition. Instead, the Type 1 diabetes discussion forum on Diabetes.co.uk was identified as being the most actively used social media included in this study, with a total of 713 posts during the 10-day period.

The sample of Twitter data contained notably fewer responses to posts (0%) than the Type 1 diabetes group on Facebook (n=463, 97%) and the Type 1 diabetes discussion forum on Diabetes.co.uk (n=672, 94%), suggesting that there is a greater degree of two-way communication between users of social networks and discussion forums than micro-blogging platforms. Twitter is designed for individual micro-blogging (broadcasting), while Facebook groups and discussion forums are designed for conversation and sharing. While Twitter posts may trigger chains of responses, this is coincidental, whereas Facebook groups are explicitly designed for this and include invited members, often mirroring offline social networks [402]. The less developed conversational structures on Twitter, make it harder for patients to read all related comments, while the Facebook and Diabetes.co.uk posts included associated comments that are easily found. When this study was conducted tweets were limited to 140 characters in length, whilst Facebook and Diabetes.co.uk provided much richer posts, both in terms of length and structured content, i.e. long chains of comments.

When analysing the data further, it was noticed that the community moderator who posted questions from anonymous members of community created all the original posts within the Facebook community. This thus created a degree of uncertainty, as the number of authors contributing to original posts is not available. This is in contrast to the Diabetes.co.uk discussion forum for Type 1 diabetes and Twitter, where any registered member of the site could generate an original post and where 37 and 62 members created an original post, respectively.

The results for the first research question in this study indicate that patients do use different social media platforms for different purposes, as are highlighted by the volume and conversation structures represented in the sample. The Type 1 diabetes group on Facebook appeared to be heavily moderated as indicated by all original posts having been authored by a single account id; therefore suggesting that prior to posting them the moderator conducted a pre-screening of messages from community members. This was weakly present in the Type 1 diabetes discussion forum on Diabetes.co.uk and absent on Twitter. Given these differences in utility, it is natural to ask how members of the Type 1 diabetes community use these social media to interact with others. Surprisingly, it was found that the Type 1 diabetes discussion forum on Diabetes.co.uk was the most actively used social media in terms of volume, whilst the Type 1 diabetes Facebook group achieved the highest percentage response rate to the original posts. This is a finding widely known within the computer science community, but not yet reflected in much of the published public health research that is available.

The results of Bales IPA reveal differences in the nature of interactions between users of these three forms of social media. Whilst the three social media are predominantly used to disseminate suggestions, opinions and information with other members, the highest percentage of posts, Twitter (n=24, 36%), Type 1 diabetes group on Facebook (n= 158, 33%) and Type 1 diabetes discussion forum on Diabetes.co.uk (n=392, 55%), represent members sharing their opinion. Noticeably fewer posts ask to receive suggestions, opinions and information from other members and the majority of these are original posts.

Several interesting differences in the nature of the posts are observed. As noted, in this sample, Twitter is mainly used to disseminate information (n=19, 29%) and opinion (n=24, 36%) and was not used for interaction. These posts were characterised by dramatisation (n=12, 18%), few friendly posts (n=12, 18%) and no indications of agreement or disagreement between members. Some examples of these tweets are shown below:

- *“33k kids in canada went #backtoschool with diabetes. it’s time to make school a better places for t1d kids”*

- *“sanofi launches mobile game for kids with type 1 diabetes in the uk”*
- *“did you know that the character elsa from the movie ”frozen” was in part inspired by a child with type 1 diabetes?”*
- *“amazing revolution - bionic pancreas which will automatically inject insulin to type 1 diabetes patients”*
- *“an open letter to teresa may advice on her type one diabetes”*

In contrast, the Type 1 Diabetes Facebook community and Type 1 Diabetes.co.uk discussion forum were considered to be very interactive, with friendly posts accounting for 47% (n=177) of Facebook posts and 46% (n=328) of posts on Diabetes.co.uk. The Type 1 diabetes group on Facebook generated a greater percentage of agreement (n=45, 12%) and disagreement (n=19, 5%) compared to the other social media. Meanwhile, the posts within the Type 1 diabetes discussion forum on Diabetes.co.uk indicate a higher percentage of tension (n=86, 12%) and unfriendly posts (n=21, 3%), particularly in relation to topics such as diet and the new treatments that are available. Given that there were 37 authors of original posts on the Diabetes.co.uk discussion forum, compared to the single account id posting members messages in the Facebook community, this suggests that there was less moderation on the discussion forum, hence enabling a greater diversity of opinion to be represented.

The three most popular discussion threads on the Type 1 diabetes Facebook community are listed below, including the number of replies to the original post in parentheses.

- *“Over/Under time again... 153 Are you over or under?”* (n=101)
- *“This may seem like an odd question but I’m more than a little curious if other T1s experience what I do. I get bit by mosquitos all the time. So much more than anyone else I know. Everyone jokes that mosquitos must love me cause my blood is so sweet. I laugh it all off cause it seems ridiculous. However, I also seem to attract bees. Kind of odd huh? I’m curious how many others experience this...if any?”* (n=63)
- *“Just want peoples advice really I’m 22yrs old been t1 diabetic since I was 6trs old I have one child but planning another but can’t get my hba1c past 8.5 and the docs won’t let me try till it’s 7 any tips on how to get it down? I’m on injections novo rapid and levemir, thanks”* (n=54)

The three most popular discussion threads on Type 1 diabetes discussion forum on Diabetes.co.uk are listed below, including the number of replies to the original post in parentheses:

- *"How highly would you recommend eating low carb? I know this is probably a silly question as it has quite an obvious answer! I love my carbs. I love pasta, rice and potatoes. In the past, I've tried to take the right amount of insulin to cover this but it's so easy to get it wrong and misjudge it - it also means I can end up taking whopping amounts of insulin! Would you recommend I reduce my carbs? It should make my diabetes easier to manage, yes? (I'm also doing Slimming World so although I've read a little about LCHF, I'm not keen to start eating loads of 'fattier' foods!)"* (n=91)
- *"New Flash Glucose Monitoring from Abbott Bloodless Testing Its arrived and heres a video for all you guys who wanted more info"* (n=76)
- *"LCHF success stories from type 1's I created this thread as a place for fellow type 1 diabetics to share their success stories on the LCHF diet. I know there is a similar thread on the low carb forum but I found that most of the responses were from type 2's, so I thought it'd be nice to have a specific place for us to share experiences and hopefully inspire and learn from each other!"* (n=67)

These results indicate that the three social media were all being used to disseminate information about the condition Type 1 diabetes. However, the Type 1 diabetes group on Facebook and the Type 1 diabetes discussion forum on Diabetes.co.uk were also being used for social interaction and peer support.

Limitations

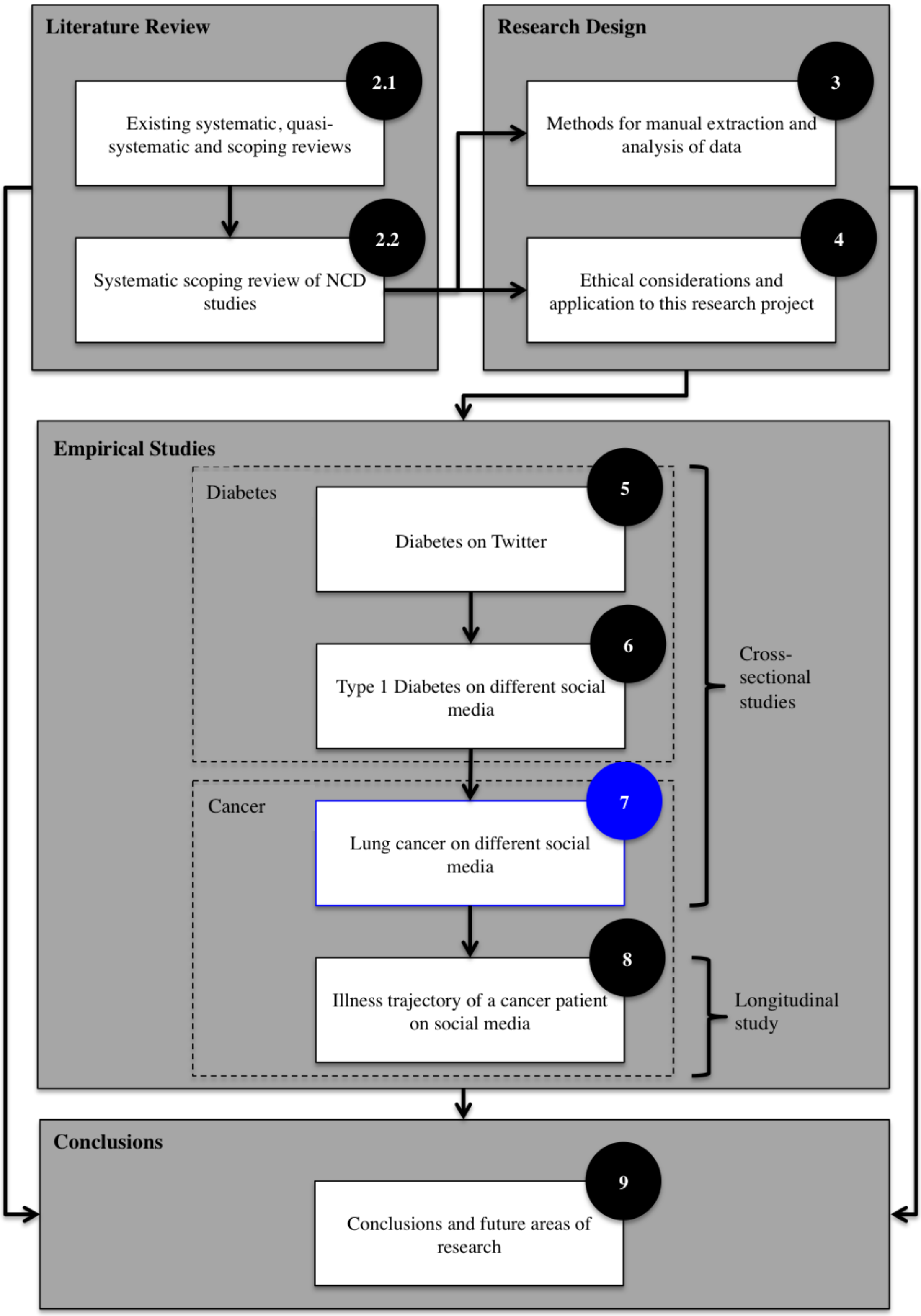
Conclusions were limited by the type and named social media platform and communities that were selected for this study, as these did not include other social networks or online communities that were available to people living with Type 1 diabetes e.g. www.TuDiabetes.org, www.patientslikeme.org, www.diabetesdaily.com or the www.WebMD.com Type 1 diabetes forum. The limitations of using Twitter as a source of data for research purposes are the same as those discussed in the pilot study (Section 5.4). Other limitations include the self-selection of contributors, as all three social media platforms require an account in order to post a message, as well as the interpretation bias, provenance and misinformation in the data.

6.5 Conclusions

Findings from this study suggest that although the three social media can be considered valid sources of information about the clinical condition of Type 1 diabetes, differences in their utility within healthcare defy the generalisations made in existing public health studies [75, 78]., where different social media are often considered equal or synonymous and therefore the findings erroneously generalised to other platforms. Type 1 diabetes patients use different social media for different purposes, with Twitter primarily used for information and opinion sharing, as opposed to making requests for information, opinion or suggestion or in expressing agreement or disagreement. Whilst Diabetes.co.uk and Facebook, by virtue of their user base, design and self moderating communities, had a greater percentage of friendly interactions, agreements and disagreements, suggesting that they are more successful in their utility for social interaction and peer support. This is a finding that has important implications for social media and their application in the context of healthcare when looking to understand how those living with a life long condition interact online.

6.6 Implications for this PhD research project

Although this study revealed interesting insights into how Type 1 diabetes patients use different social media platforms for different purposes and which social media are successful at encouraging social interaction and support for patients, it would be erroneous to generalise these findings to represent all NCDs. These implications, and whether the utility of different social media varies depending on the type of NCD under investigation, warranted further research and empirical evidence. As highlighted during the systematic scoping review (Section 2.2), there were few studies that source data from social media, which investigated lung cancer and hence this gap formed the basis of the third empirical study conducted as part of this PhD research project and described in Chapter 7.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

7. Are the differences in social media usage and associated data types also seen in people affected by lung cancer?

As highlighted in the results of the systematic scoping review (Section 2.2), lung cancer is under-represented in the existing literature. Expanding on the findings from the Type 1 diabetes study described in Chapter 6, the third empirical study conducted as part of this PhD research project aimed at addressing RQ6 and focuses on the different utility and affordances offered by different social media to people affected by lung cancer. The outcome of this study was published in the *npj Digital Medicine* journal (Appendix A.1.3).

7.1 Context and aim of the study

Lung cancer is the most common cancer globally, with approximately 2.09 million cases every year [403]. There are two main types of primary lung cancer: small cell lung cancer (SCLC), affecting 15% of those diagnosed in the United States, and non-small cell lung cancer (NSCLC), affecting 85% [404]. The severity of this condition varies between Stage 1 and Stage 4, depending on the size of the tumour and whether it has spread. These stages can influence survival rates, with fewer than 6% of patients living more than five years after diagnosis with Stage 3. Treatment options include but are not limited to surgery, chemotherapy, radiotherapy and laser therapy and are dependent on the individual. Causes of lung cancer include smoking, passive smoking, exposure to radon and asbestos, air pollution, low immunity and family history, to name but a few [405]. While smoking is a known risk factor, 10-15% of people who develop lung cancer are never-smokers and its cause cannot be definitively associated with established environmental risk factors [406]. Much research has been conducted into the stigma associated with lung cancer [407-409] and has shown that it is considered to be more highly stigmatised than other cancers [410, 411] due to self-blame and its causal attribution to smoking [412, 413]. As such, those diagnosed with lung cancer are encouraged to seek support through support groups and online communities in order to reduce the likelihood of depression [414].

Hamm et al.’s scoping review of studies involving social media use by patients and caregivers [69] reported that discussion forums, online support groups, social networking sites and micro-blogs dominate the research literature, with 11% of the identified studies focusing on cancer. Although relevant research is fragmented and currently lacking in substantive empirical evidence, existing studies suggest that social media can be used to provide social, emotional or experiential support in chronic disease management and are likely to improve patient care [86, 415]. Psychological support was revealed to be present in the majority of tweets by cancer patients [416], whilst a narrative synthesis of cancer patient blogs indicated that users share their diagnosis and treatment journeys online as a means of describing their experiences of health services, informing their health behaviour and in maintaining relationships with others [379]. Lung cancer is the second most prevalent cancer discussed on Twitter, after breast cancer [417], and research has revealed that the majority of relevant tweets focus on treatment and the use of pharmaceutical and research interventions, followed by awareness-raising and prevention/risks [418].

Health researchers are increasingly using data sourced from social media to understand how members of patient communities interact with each other regarding specific conditions [75, 76]. Single platforms, such as Twitter and Facebook or condition-specific online communities, have dominated previous research, as highlighted in earlier chapters.

The empirical study described in Chapter 6 is expanded upon here, by exploring the types of interaction and support demonstrated on different social media by people affected by lung cancer, with specific reference to lung cancer hashtags on Twitter [8], the Lung Cancer Support Group on Facebook [419] and the lung cancer discussion forum on Macmillan.org.uk [420]. In doing so, the following research questions were considered:

i. Do people affected by lung cancer use different social media in different ways?
ii. Which social media are most successful at encouraging social interaction and support for people affected by lung cancer?

7.2 Methods

Similar to the empirical study described in Chapter 6, three social media were selected that have different characteristics. Twitter and Facebook are general platforms for information sharing and social networking. In the case of Twitter, communities and topics are often collectively organised around a set of hashtags, which can then be searched for to understand particular issues, as in this study. Users of Facebook can set up discussion groups focused on certain topics, including lung cancer. Specialist organisations, such as Macmillan Cancer Support, may also set up condition-specific online discussion forums, where users can contribute content and others can comment. Users of social media must create an online profile and account ID when registering for these platforms. In doing so, it is within their gift to disclose personal information if they so choose, despite it being discouraged for safety reasons. This variability results in some users being more easily identifiable as a patient, carer or family member than others. Given this incompleteness of profile data, the analysis treats patients and their carers or family members as one group of people affected by lung cancer.

To compare lung cancer-related interactions on different social media, a multi-stage approach was applied; a) extracting and screening posts appearing in each social media platform, b) classifying posts using Bales' IPA [361], c) categorising posts according to the four functions of social support and d) analysing the 100 most frequent keywords to generate semantic 'word clouds' (using Wordle [421]) to visualise the frequency of terms used in posts associated with each form of social support.

7.2.1 Data sources and screening

Twitter. A sample of 3,000 tweets was extracted using Symplur's Transcript and Analytics tool in December 2017, using the #LungCancer and #LCSM hashtags [422, 423]. These hashtags were selected due to their specific relevance to the condition and as a means of reducing the amount of 'noise' presented in the data. The sample of Twitter data extracted and analysed in this study does not encompass the universe of all Twitter data available. Due to the limitations of the Transcript and Analytics tool, the sample size is limited to 1,500 tweets per hashtag and the timeframe for the extracted tweets was between 11:55PM on 30 September 2017 and 01:00AM on 1 November 2017. While the transparency of Symplur's search and sampling algorithms has been criticised, due to its propriety nature and hence the lack of publically available information and clarity of its effectiveness, its use in over 280 published research articles [424] provided justification for its use in this study. The data extracted included the Twitter account ID and the text in the tweet. Re-tweets and tweets that were not in English were excluded from the categorisation stage. URL links and images included in the tweets were not captured or reviewed during the screening.

Facebook. Using the Facebook search functionality, the search term 'lung cancer' identified the largest lung cancer community available on Facebook. Known as the Lung Cancer Support Group, this community was established in 2015 and is a closed group for lung cancer patients, survivors, caregivers and loved ones. Closed groups are members-only groups, where the group's existence is visible to anyone with a Facebook account. Membership of the group, however, is granted through the group administrators. Access to the Lung Cancer Support Group was granted via correspondence sent to the group administrators explaining the intent behind our request. As of 30 December 2017, it had 7,975 members and, on this same date, all wall posts and replies that were posted between 1 and 31 October 2017 were identified by using the search functionality and stipulating the posts could be posted by ‘anyone’, ‘anywhere’ and in ‘October 2017’. This sample of original posts and any associated replies posted in October 2017, along with the author and the date of post, were manually extracted for further analysis. Posts that were not in English were excluded from the categorisation stage. URL links and images included in the posts were not captured or reviewed during the screening.

Macmillan.org.uk. Macmillan Cancer Support is a UK-based charity that was founded in 1911 and provides specialist health care, information and financial and emotional support to people affected by cancer. It has an online community of over 100,000 members [425], spanning all forms of the condition. The lung cancer discussion forum on Macmillan.org.uk was identified through the search functionality of the online community homepage and the list of discussion threads was then filtered based on the start date 1 October 2017. All original posts and replies posted between 1 and 31 October 2017 were identified and manually extracted. The total sample of 266 posts was included in the categorisation stage.

7.2.2 Application of Bales’ IPA and the social support taxonomy

The same sample of lung cancer posts included in the categorisation stage was reviewed against the twelve categories of group interaction defined in Bales’ IPA (Figure 33), as well as the four functions of social support (i.e. emotional, informational, instrumental, companionship). This was done in order to provide deeper understanding of the type of support garnered by users of social media as well as the interaction between users of social media. Each post was considered to be a single unit of interaction and the categories were not considered to be mutually exclusive when applied to the sample of posts.

7.2.3 Analysis of key words

The content of lung cancer posts meeting each category of social support was analysed by inputting the data into the text visualisation tool Wordle [421]. The resulting semantic word clouds map the frequency and co-occurrence of different terms appearing in a corpus of text and can be used by researchers to compare the topics and sentiment appearing in different text [426]. The word clouds produced in this study represent the top 100 words with the highest frequency of occurrence, in alphabetical order. Hashtags and terms such as ’lung’, ‘cancer’ and ’lungcancer’ were removed, as these had already been used to screen the social media posts for inclusion in the study.

7.3 Results

Table 21 shows the number of lung cancer-related posts extracted from each of the three social media, the total number of people contributing to each platform, the number of replies and the number of relevant English language posts included in the categorisation stage. Posts that were excluded were not in the English language. The Twitter hashtags #LCSM and #LungCancer were found to have the highest absolute number of posts (3,000 posts over the one-month period), followed by the Lung Cancer Support Group on Facebook (2,644 posts) and the lung cancer discussion forum on Macmillan.org.uk (266 posts).

Table 21. Summary of social media posts about lung cancer from October 2017

Source	Number of unique authors	Total sample size	Number of original posts	Number of replies	Number of secondary replies	Number of posts included in IPA categorisation stage (%)
Twitter #LCSM and #LungCancer	1,056	3,000	3,000	0	0	2897 (97%)
Facebook Lung Cancer Support Group	844	2,644	51	1,659	934	2597 (98%)
Macmillan.org.uk lung cancer discussion forum	96	266	51	215	0	266 (100%)

The Twitter hashtags had the highest number of unique authors (1,056), compared to the Lung Cancer Support Group on Facebook (844) and the lung cancer discussion forum on Macmillan.org.uk (96). At 2593, the Lung Cancer Support Group on Facebook had the highest number of replies compared to the Macmillan lung cancer discussion forum (215 replies) and the Twitter hashtags (0 replies).

Table 22 and Figure 34 show the frequency of posts fitting each of the 12 categories developed by Bales [361] (Figure 33) for classifying the type (not content) of interactions taking place in groups, known as IPA. While all three social media were mainly used to post suggestions, opinions and information, *information sharing* was the most common use, representing 64% (n=1849) of posts bearing the Twitter lung cancer hashtags, 58% (n=154) of posts to Macmillan's lung cancer discussion forum and 43% (n=1117) of posts to the Lung Cancer Support Group on Facebook. Noticeably fewer posts sought suggestions, opinions and information from other members, although this varied across platforms (5%, n=158 of Twitter lung cancer hashtag posts, 7%, n=195 of posts in the Lung Cancer Support Group on Facebook and 28%, n=74 of the posts on Macmillan's lung cancer discussion forum). There were also more posts classified as *friendly, unfriendly, tension release and showing tension* in the sample from Macmillan's lung cancer discussion forum (57%, n=151 friendly, 6%, n=16 shows tension, 6%, n=15 tension release, 1%, n=3 unfriendly) compared to the Lung Cancer Support Group on Facebook (38%, n=973 friendly, 2%, n=43 shows tension, 0.7%, n=17 unfriendly, 0.4%, n=10 tension release) and the Twitter lung cancer hashtags (11%, n=326 friendly, 0.6%, n=18 unfriendly, 0.4%, n=11 shows tension release, 0.2%, n=6 tension release), suggesting a greater degree of sentiment expressed in the Macmillan discussion forum. To provide transparency and increase the reproducibility of our analysis, examples of posts from each category and social media platform are provided in Table 23.

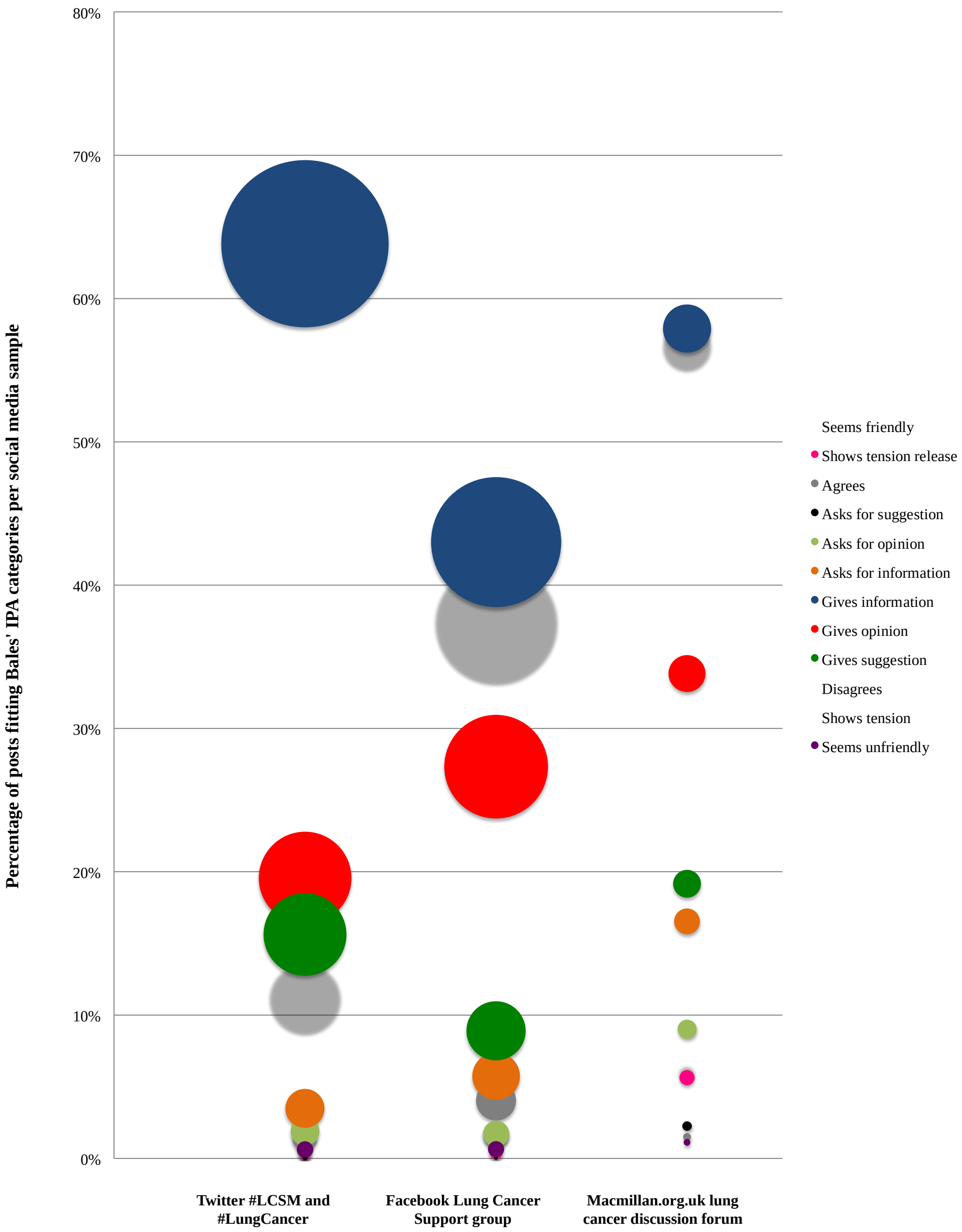
.

Table 22. Application of Bales' IPA to the lung cancer posts

Source	Total number of posts analysed	Seems friendly	Shows tension release	Agrees	Asks for suggestion	Asks for opinion	Asks for information	Gives information	Gives opinion	Gives suggestion	Disagrees	Shows tension	Seems unfriendly
Twitter #LCSM and #LungCancer	2897	326 (11.3%)	6 (0.2%)	39 (1.3%)	2 (0.1%)	55 (1.9%)	101 (3.5%)	1849 (63.9%)	567 (19.6%)	453 (15.6%)	1 (0.03%)	11 (0.4%)	18 (0.6%)
Facebook Lung Cancer Support Group	2597	973 (37.5%)	10 (0.4%)	105 (4.0%)	1 (0.04%)	45 (1.7%)	149 (5.7%)	1117 (43.0%)	710 (27.3%)	231 (8.9%)	0 (0%)	43 (1.7%)	17 (0.7%)
Macmillan.org.uk lung cancer discussion forum	266	151 (56.8%)	15 (5.6%)	4 (1.4%)	6 (2.3%)	24 (9.0%)	44 (16.5%)	154 (57.9%)	90 (33.8%)	51 (19.2%)	0 (0%)	16 (6.0%)	3 (1.1%)

The categories were not considered to be mutually exclusive

Figure 34. Percentage of posts in the corpus of data from the Twitter lung cancer hashtags, the Lung Cancer Support Group on Facebook and the lung cancer discussion forum on Macmillian.org.uk, and their fit to Bales' IPA categories



The size of each circle represents the percentage of posts associated with each of Bales' IPA categories. Each colour represents a different category: light blue for 'seems friendly', fuchsia for 'shows tension release', grey for 'agrees', black for 'asks for suggestion', light green for 'asks for opinion', orange for 'asks for information', dark blue for 'gives information', red for 'gives opinion, dark green for 'gives suggestion', lilac for 'disagrees', turquoise for 'shows tension' and purple for 'seems unfriendly'

Table 23. Examples of the different categories of posts from the three social media

	Twitter #LCSM and #LungCancer	Facebook Lung Cancer Support Group	Macmillan.org.uk lung cancer discussion forum
Seems friendly	‘I'm so saddened & disappointed. My condolences to her loving family & the lung cancer community’ ‘Thanks everyone who walked today’	‘Sending love and prayers’ ‘Good luck next week’	‘I’m so so sorry for your loss. Your mom fought and lost but she is now in a better place. I don’t know what to say other than I’m thinking of you and her. Please take care and your time to grieve’
Shows tension release	‘US #1 cancer killer is Lung Cancer:433 die daily,160K die yearly! Lungs are sexy, too!’	‘Nobody is getting out of here alive. Life is a 100% fatal disease.’	‘My dad’s taste has completely changed! He’s like a pregnant woman!’
Agrees	‘Agreed! Either light up the White House for every cancer or don't light it up at all! #WhyIsThisPink’	‘I agree with you.....hospitals are scary’ ‘I totally agree. Family and friends may be supportive but they don’t understand what we’re going through and feeling like everyone on here’	‘My grandad has been having the sweats! Definitely agree. They should make these symptoms more aware. Especially considering how aggressive this cancer is.’
Asks for suggestion	‘How might doctors and other healthcare providers better integrate #yogatherapy into their offerings?’	‘Where and how do all of you find your inner strength? I'm so afraid. Please share what helps you.’	‘My dad has advanced lung cancer. …. Does anyone have any suggestions for good support pillows that helped them or their loved ones?
Asks for opinion	‘Which lung cancer group(s) do you favor? Why do you like them?’	‘Is there hope for Lung cancer? Need some warriors here to post some good news for a change ....’	‘If the 4 chemo sessions I'm going to get are only going to give me a short time extra and in that short time I'm going to feel dreadful, then what's the point? Does that sound a bit defeatist?’
Asks for information	‘Which Genetic Tests Are Important for Metastatic Lung Cancer Treatment?’	'Does it mean that it's not advanced NSCLC since my dad tested negative for the PD-L1?’	‘Found out today that if the disease progresses they would consider second line treatments such as Docetaxel or +/- Neratinib. Has anyone any experience of these?’
Gives information	‘#BreakingNews: FDA approves new dosing option for ALUNBRIG (brigatinib) in the US’	‘Keytruda did not work for me because I have a ALK mutation in my cancer. I went to different oncologist and she put me on a targeted therapy for that mutation and my tumors are shrinking.’	‘I am brand new and so thankful that I found your site. I went to the ER yesterday for abdominal pain and they did an abdominal CT scan. They found 2 small nodules (2 mm and 3 mm) in one of my lungs. It says 2/3 „ stable densities suggesting benignity.’
Gives opinion	‘Let's not forget that #lungcancer takes more lives than the next four cancers put together. More research clearly needed’	‘It’s awful but it gets better when the radiation stops. Just work hard on staying hydrated and eating what you can.’	‘Things will get easier when you have a treatment plan and can get your head round it and see the way forward.’
Gives suggestion	‘For those who've lost someone to lung cancer. Join me at the Life & Breath Rally.’	‘Take it day by day and test by test….Stay away from the internet searches and Dr. Google.’	‘You can carry out a search within the Lung group typing some words into the search bar and it should list all discussions containing that text.’ ‘Have you been allocated a Macmillan nurse? If not speak to the oncology department.’
Disagrees	‘This seems wrong’		
Shows tension	‘Damn it! xxx died’ ‘This f*kin' sucks. She so wanted to be in Oahu with her family. I hope she was.’	‘we need to STOP the Stigma and not worry about who smokes, smoked or never smoked…it causes so much upset and hard feelings because so many people still "assume" that you had to smoke to get Lung Cancer’	‘The consultant didn't want to put a time frame on her life expectancy … how the hell do you get through this? How do I explain to my 4 year old when the time comes that grandma isn't coming over again?’

	Twitter #LCSM and #LungCancer	**Facebook Lung Cancer Support Group**	**Macmillan.org.uk lung cancer discussion forum**
Seems unfriendly	‘… there is no need to be rude. I am a cancer survivor. I don't think I brought it on myself & I changed my career to advocate for those who feel their voices are lost…’	‘SPAM!...You are in violation of the GROUP RULES!’ ‘Comments like that are so ignorant’	‘My original post was supposed to provide my Dad with some positive stories and tips…I dont think i can share this thread with him after the last post. Im sympathetic to your situation but just ask you be the same to mine’

Table 24. Frequency of the four different functions of social support on each of the three social media

Source	Total number of posts analysed	Emotional support	Instrumental support	Informational support	Companionship support	Not considered social support
Twitter #LCSM and #LungCancer	2897	168 (5.8%)	9 (0.3%)	842 (29.1%)	686 (23.7%)	1307 (45.1%)
Facebook Lung Cancer Support Group	2597	1333 (51.3%)	0 (0%)	1421 (54.7%)	71 (2.7%)	0 (0%)
Macmillan.org.uk lung cancer discussion forum	266	178 (66.9%)	0 (0%)	174 (65.4%)	33 (12.4%)	0 (0%)

The categories were not considered to be mutually exclusive.

After excluding promotional or news posts (45% of the Twitter sample only), data from the three platforms was analysed against the four-dimensional typology of social support [427-429]. The four social support types are Emotional (offering empathy, concern, affection, love, trust, acceptance, intimacy, encouragement or caring), Instrumental (provision of financial assistance, material goods, services or tangible aid), Informational (provision of advice, guidance, suggestions or useful information to someone) and Companionship (design to give a sense of belonging). The results are summarised in Table 24. These highlight differences between the three social media, with informational support having a higher proportion on the discussion forum on Macmillan.org.uk (65%, n=174) and Lung Cancer Support Group on Facebook (55%, n=1421) compared to the Twitter hashtags (29%, n=842). Emotional support was also most evident in the lung cancer discussion forum on Macmillan.org.uk (67%, n=178), compared to Lung Cancer Support Group on Facebook (51%, n=1333) and the Twitter hashtags (6%, n=168).

The frequency and co-occurrence of keywords associated with posts in each social support category from all three forms of social media were mapped into the semantic ‘word clouds’ shown in Figure 35. No word cloud was produced for ‘Instrumental’ support, due to the small number of posts in this category. The preliminary themes characterising posts falling into each of the remaining three social support categories were identified through observational means and are shown in the right-hand column. Emotional support is represented by qualitative themes such as *spirituality, grief, family and positive sentiment*, whilst informational support is identified by terms related of the *diagnosis and treatment* of the condition. Key words relevant to *community and advocacy* are evident in the category of companionship support.

Figure 35. Semantic word clouds visualising the frequency of words by social support function for lung cancer

Function of social support	**Semantic word cloud**	**Themes identified**
Emotional support		Spiritual – faith, prayers, bless, hope, god Grief – sorry, loss Family – dad, mom Positive sentiment – thanks, hugs
Informational support		Diagnosis – symptoms, questions, scan, oncologist, stage, spread, brain Treatment – chemo, radiation, therapy, surgery, pain, Time - today, days, weeks, months, years
Companionship support		Community – join, chat (date and time), welcome, group, team, social, support, help, everyone Advocacy – rally (location and date), lung cancer awareness month, awareness

7.4 Discussion

This descriptive analysis, encompassing a total of 5,910 relevant messages posted on Twitter using the hashtags #LCSM and #LungCancer, the Lung Cancer Support Group on Facebook and the Macmillian.org.uk lung cancer discussion forum. Across the three forms of social media, the absolute number of postings in the one-month observation period was greatest for the Twitter hashtags, followed by the Lung Cancer Support Group on Facebook and then the lung cancer discussion forum on Macmillan.org.uk. However, these raw numbers say little about their value for users and hence the further analysis was done that identified variations in the use of different social media by people affected by lung cancer. These variations include the nature of interactions within these online communities and the type of social support represented.

As noted in Table 22, the Bales' IPA highlighted similarities and differences in the type of interactions found on each of the three social media platforms. All were predominantly used for information sharing, with the lung cancer discussion forum on Macmillan.org.uk also showing the most posts expressing sentiment (e.g. friendliness, tension).

There are several possible explanations for these differences. One relates to the digital architectures of the platforms – defined as ‘the technical protocols that enable, constrain and shape user behaviour in a virtual space’ [430: pg471]. Twitter is designed for individual micro-blogging (broadcasting), while Facebook groups and discussion forums are designed for conversation and sharing. While Twitter posts may trigger chains of responses, this is coincidental, whereas Facebook groups are explicitly designed for this and include invited members, often mirroring offline social networks [402]. In this study, high levels of two-way communication were seen in the Lung Cancer Support Group on Facebook (99.8%) and lung cancer discussion forum on Macmillan.org.uk (81%), while the sample of Twitter posts contained only original tweets and re-tweets.

A preliminary observation that warrants further investigation relates to the types of account holders posting to the different social media platforms. In previous analyses of health-related Twitter narratives, 25% of verified accounts belonged to journalists [402], 40% to companies and brand accounts and 15%-20% were bots [397]. In this study, it appears that organisations (e.g. pharmaceutical companies, charities), patient advocacy groups, research institutions and news outlets use Twitter to disseminate health information to the general public. In contrast, the Lung Cancer Support Group on Facebook and the discussion forum on Macmillan.org.uk were chiefly designed for people affected by lung cancer and due to the presence of community moderators, function as online support groups and enable a greater degree of interaction.

Community moderators or administrators play an influential role in online communities [431, 432], as was the case in the closed (members only) Lung Cancer Support Group on Facebook and the lung cancer discussion forum on Macmillan.org.uk, where their presence created greater understanding of the community rules and expectations. In the case of the Lung Cancer Support Group on Facebook, group members were asked to agree to a code of conduct, requiring them to respect others, avoid foul language, focus on lung cancer, not seek medical advice, refrain from advertising or fund-raising and never block an administrator (i.e. moderator) of the group, or face removal. As well as providing a ‘safe’ place for people to share information and experiences, moderators of the Macmillan.org.uk lung cancer discussion forum responded to posts that had not yet received a comment from other community members.

Examples of this include *'sorry to hear your news, and I'm also sorry you've not had a reply yet*'. In responding to this post, the moderator was able to increase the visibility of the post in an attempt to garner a response. In contrast, moderation of content posted on Twitter is absent, except during the one-hour pre-scheduled #LCSM group discussion that takes place every two weeks [433]. The moderator in these cases is there to greet participants and facilitate a discussion based on predefined questions (e.g. *'I will announce four topic questions (T1 T2 etc). Pls label your answers with T1, T2, etc to make transcript easier to follow #lcsm').*

Findings also revealed that posts by any user that are similar to an online survey would generate the most responses on Facebook, as was also the case in a previous diabetes study described in Chapter 6 [434]. Examples of this include: '*how old was everyone when they were first diagnosed?'* (609 responses), *'this may be a really dumb question: can someone survive lung cancer?*' (228 responses) and '*anyone got a husband or wife that won't give up smoking?'* (73 responses).

In summary, whilst the Lung Cancer Support Group on Facebook and lung cancer discussion forum on Macmillan.org.uk were moderated, this is limited on Twitter to pre-scheduled group discussions. Twitter was found to be the most actively used social media in terms of volume of posts, whilst Facebook achieved the highest percentage response rate and interaction. The differences in digital architecture, in turn, contribute to the variations in social interaction and support for people affected by lung cancer.

Several interesting differences in the social interaction represented in the posts on these three social media are observed. As previously noted, the lung cancer hashtags on Twitter was mainly used to disseminate information (64%) and opinion (20%) in line with research describing Twitter as a mass communication and broadcasting tool [435], with the majority of active narratives involving two or fewer users [436]. Findings from this study are similar in nature to that of the Type 1 diabetes comparative study described in Chapter 6 [434].

The positive correlation between social support and health is widely accepted amongst the public health and psychology research communities [437]. In addition to types of functional interaction shown in the posts, this analysis revealed differences in the types of social support that were being sought and offered by users of different social media. All content posted in the Macmillan.org.uk lung cancer discussion forum and in the Lung Cancer Support Group on Facebook was associated with at least one of the functions of social support, whilst 44% of tweets were not and, in most cases, can be considered as either news reports (e.g. *'FDA Approves Higher-Dose Tablet of Brigatinib for NSCLC'*), promotional messages (e.g. '*Get your #lungcancer swag for #LCAM17 this Nov. T-shirts, tank tops, sunglasses, bracelets and more!'*) or advocacy-related (e.g. *'#433aday Lung cancer kills 433 Americans a day. We need better funding for research now')* in content.

Informational support, as already noted, is the provision of information aimed at supporting a member or members of the social network, often in response to a statement of distress or a request for help. While this overlaps with Bales' IPA categories of 'information giving', they are not synonymous. The lung cancer discussion forum on Macmillan.org.uk produced the highest percentage of informational support posts (66%), followed by the Lung Cancer Support Group on Facebook (54%), the #LungCancer on Twitter (41%) and the #LCSM on Twitter (15%). Posts such as *'Looking for advice. Mum is in later stages of stage 4 lung cancer....In the last 4 days mum has stopped eating and can only get a small amount of fluids in her...Not sure what to do or what to expect now'* and *'can radio be used after Keytruda? Can it be keytruda + radio?'* illustrate the type of informational support sought. Posts classified as informational support are represented in words related to the diagnosis, treatment and progression of the condition over time, a finding which accords somewhat with findings from Tsuya et al.'s study where keywords within the tweets reflected the use of Twitter as a tool for sharing medical information amongst cancer patients [379].

The lung cancer discussion forum on Macmillan.org.uk produced the highest percentage of emotional support posts (67%, n=178), followed by the Lung Cancer Support Group on Facebook (51%, n=1333) and the lung cancer hashtags on Twitter (6%, n=168), contradicting findings from previous research that revealed that the majority of tweets posted by cancer patients focused on psychological support [438]. Emotional support is represented in posts such as *'I am so sorry to hear this...you are in my thoughts'* and *'It really is so hard...I'm full of hurt and anger. Just seems so unfair'*. Words associated with emotional support include spiritual and religious terms (e.g. faith, god, hope, prayers), grief (e.g. sorry, loss), family and positive sentiment (e.g. thanks, hugs).

Given the severity of lung cancer, its treatment and the life-threatening nature of the condition, it is not surprising that spiritual and existential beliefs are represented in the social media data [439], Lung Cancer Support Group on Facebook (15%), lung cancer discussion forum on Macmillan.org.uk (3%) and lung cancer hashtags on Twitter (1%). Identified as one of the modifiable dimensions of the patient experience, much research has been conducted into the role spirituality and faith plays in the illness trajectories of lung cancer patients [440-442]. Its manifestation in social media, however, is relatively new [443-445], with little empirical evidence of how social media platforms differ in this context. Findings from this study suggest that spirituality is noticeably more prevalent on Facebook rather than Twitter and further research is warranted to test this hypothesis.

Surprisingly, the Twitter lung cancer hashtags indicated the highest percentage of companionship support (24%) compared to the lung cancer discussion forum on Macmillan.org.uk (13%) and Lung Cancer Support Group on Facebook (3%). Words associated with this form of social support include those related to joining online conversations and physical events, manifesting themself in posts such as *'Sending so much love to all of you. Love my tribe. Let's do this'*, *'Anyone can do something…A handful of us started #LCSM'* and *'Join us on November 2 in Washington DC'*. The higher percentage of companionship support posts on Twitter may be attributed to the use of the #LCSM as a means of forming topic communities where large groups of users, who do not need to be connected through existing 'follower' networks, can interact within the constraints of Twitter's digital architecture [446]. The #LCSM hashtag stands for 'Lung Cancer Social Media' and was created with the intention to 'unite patients, caregivers, advocates, healthcare providers and researchers to discuss ways to improve lung cancer diagnosis, treatment, research, patient outcomes, caregiving, information sharing and public support' [422]. It connects those participating in the pre-scheduled online discussion and is supported by posts such as *'If anyone is just lurking tonight, please blank tweet the hashtag #LCSM so we know you're out there. We're a friendly bunch'* and *'Remember to include #lcsm in your tweets...'*. This suggests that, through the use of hashtags and followers [447], Twitter can be used to provide a source of social and community support to people affected by lung cancer in knowing that they are not alone and in building self-esteem, confidence and social validation [448].

Instrumental support, such as the provision of financial assistance, material goods, services or tangible aid, was not present in any of the posts in the Lung Cancer Support Group on Facebook or the lung cancer discussion forum on Macmillan.org.uk. It was, however, present on Twitter (0.3%, n=9), as users, often from the USA, requested financial support to fund their treatment *'Help me complete #LungCancer treatment #Donate #crowdfund Please retweet!'*

Other topics that are notably scarce in the samples include stigmatisation and trolling. Although health-related stigmatisation on social media has been reported in other research [449] and in specific about lung cancer [450], here it was discussed in only 2% of Twitter hashtags, 1% of posts to the Lung Cancer Support Group on Facebook and not at all in the lung cancer discussion forum on Macmillan.org.uk (0%). Examples showing how it was manifested include: '*A lot of people with cancer are afraid to talk about it, especially Lung Cancer, since some will just assume you smoke and you did it to yourself'* (Facebook user), *'Lung cancer is the biggest killer yet there is no education on it. I do believe it is the stigma of smoking that is associated with it...it is the first thing people say to me'* (Facebook user) and *'We're fighting stigma that holds lung cancer back from broad public sympathy despite being biggest cancer killer'* (Twitter user).

Trolling is defined as 'the practice of behaving in a deceptive, destructive, or disruptive manner in a social setting on the internet with no apparent instrumental purpose' [451: pg97]. Unfriendly social interactions were represented in less than 1% of posts in the corpus of data from the three social media platforms, contrasting with a recent analysis showing that 24% of online trolling incidents were associated with health-related topics [452].

Limitations

Limitations in this study reflect the self-selective nature of the contributors, as a social media account is required in order to author a post. Given the 'digital health divide' [453] and recent statistics suggesting that most people diagnosed with lung cancer are 65 or older [454], the sample does not necessarily represent the population of people diagnosed and living with this condition. This study did not go so far as to investigate the authors of the posts to determine their provenance. Analyses of social media data have their own limitations due to the presence of misinformation and the lack of provenance of account holders [75]. This can impact the validity of results and conclusions drawn from studies aimed at investigating the patient narrative and value of social media for online communities, a topic that has recently attracted media attention due to use of bot factories to boost followers and the creation of 'fake news' [455]. In addition, the sample of data was extracted during one month of the year, therefore the cross-sectional nature of the study did not take into account variations in activity that may have occurred over time, which can be useful for examining evolving narratives during cancer progression [447]. While the data from the Lung Cancer Support Group on Twitter and the discussion forum on Macmillan.org.uk were extracted manually, the Symplur Transcript and Analytics tool was used to extract the Twitter data, making the sampling less easy to verify due to the lack of transparency of the algorithms and sampling methods used. Other automated social media mining techniques and natural language processing tools are available; however, in some cases, these require software and data access licenses and can vary in their accuracy and effectiveness due to the scope and quality of data available and the types of social media for which they are suited. In identifying the preliminary themes associated with the different functions of social support a rigourous technique for thematic analysis was not applied and instead themes were observational in nature [456].

7.5 Conclusions

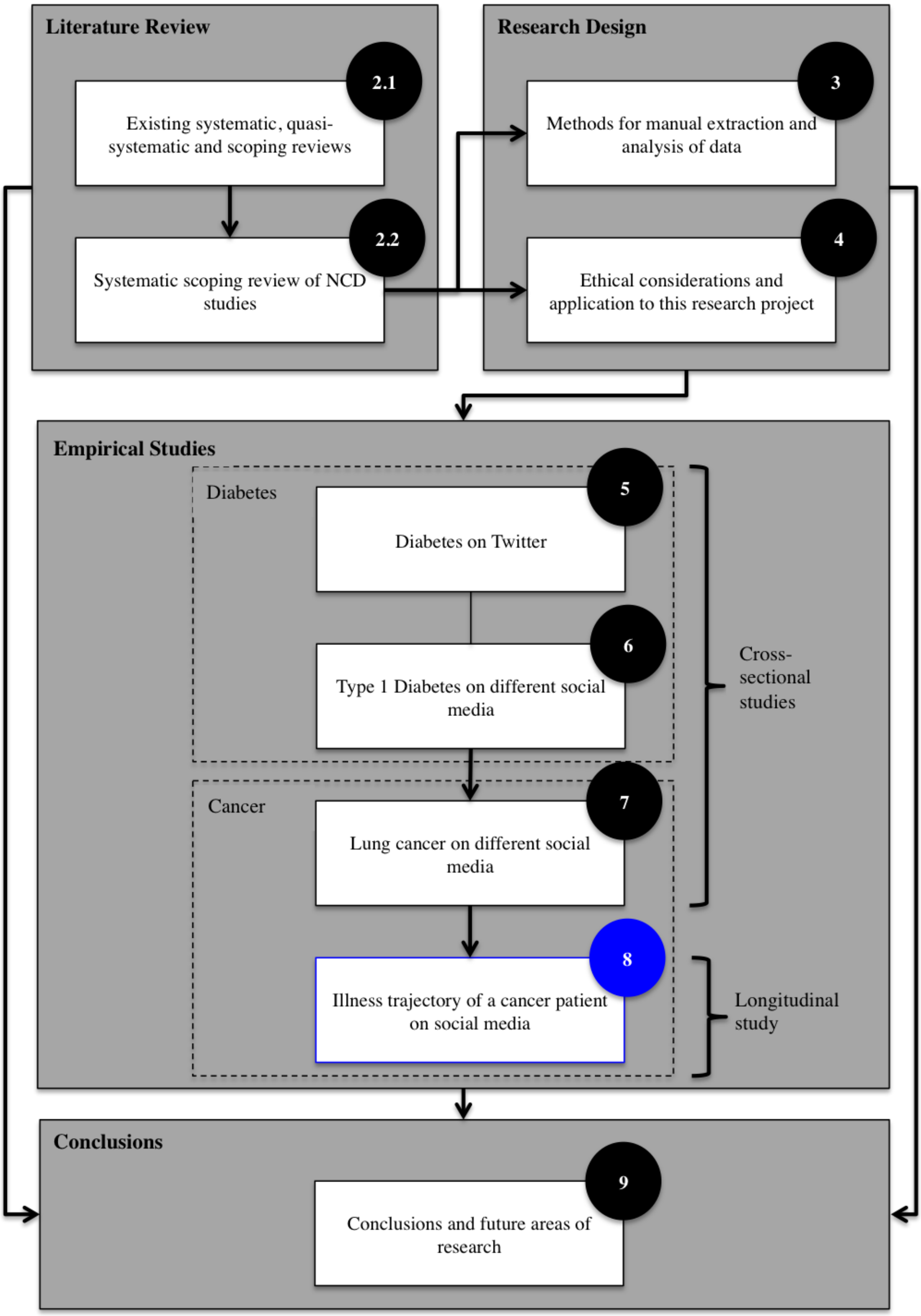
In conclusion, findings from the systematic analysis of comparable lung cancer posts on three social media indicate that, although all three were being used to disseminate information about lung cancer, the Lung Cancer Support Group on Facebook and lung cancer discussion forum on Macmillan.org.uk, by virtue of their digital architecture, user base and self-moderating communities, were more successful in their utility for social interaction and emotional and informational social support. While the sample derived from Twitter hashtags contained fewer posts related to social support across the four categories, posts tagged #LCSM showed the greatest degree of companionship support, revealing how the affordances of this platform can be shaped by its users through the use of a community hashtag. Further analysis also revealed an unanticipated sub-category of *spiritual support*, which featured uniquely in the Lung Cancer Support Group on Facebook and warrant additional research, as well as limited evidence of resentment about the comparative *stigmatisation* of lung cancer compared to other types of cancer [412].

These findings provide tentative insights into the social and supportive value of different social media, and show how interactions may be shaped both by the configuration and moderation of the platforms and by users self-organising around groups or hashtags. They suggest that health care providers and policy makers wishing to provide supportive interventions via social media, including the use of social media to discuss stigma openly [457], should prioritise community-based forums over general social media broadcasting. Likewise, academics and public health analysts wishing to study lung cancer via social media should carefully consider the types of data that are likely to appear on different platforms and their suitability for answering their research questions (e.g. whether through depth or volume). Importantly, the results also provide empirical evidence that people affected by lung cancer and those supporting them (e.g. healthcare teams, family, carers, clergy) should consider online communities as an additional source of social support during times of crisis.

While hand-searching and content analysis has proven effective in identifying and comparing the key types and expressions of social support for lung cancer manifested on these social media platforms, further research is needed to unpick, replicate and extend these findings with larger samples of data. Automated data mining and natural language processing techniques enable the capture and analysis of much larger volumes of data across multiple social media, offering the potential to create a greater degree of precision [458] when combined with the appropriate qualitative analysis. Nevertheless, it is essential that third party users, including healthcare providers, acknowledge the sensitivity of users' data, albeit data that has been voluntarily placed in the public domain by users who have willingly accepted the terms and conditions of the social media platform's use, since current research ethics guidelines governing the use of aggregated social media data remain open to interpretation as discussed in Chapter 4 [459].

7.6 Implications for this PhD research project

Findings from this study contribute to addressing the gap in empirical studies about lung cancer that was identified in the systematic scoping review (Section 2.2). Each of the previous three empirical studies described in Chapters 5-7 are cross-sectional in design and analyse a corpus of social media data posted by multiple account holders during a pre-defined period of time. Patients living with an NCD are, however, likely to share their experiences online at different intervals over a pre-longed period of time. In 2013, Love published a study that used the theory of illness trajectories as a framework to explore how work associated with cancer impacts, and is impacted by, nutrition and exercise concerns [170]. However, it did not go so far as to analyse the trajectory of a single patient over a pre-longed time. Hence, a longitudinal study was needed to determine how the experiences of patients living and dying of cancer manifest themselves in the social media narrative.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

8. Can characteristic illness trajectories be seen in a cancer patient's digital narrative and what insights can be gained to inform palliative care services?

The final study conducted as part of this PhD research project aimed at addressing RQ7 is retrospective and longitudinal in nature, analysing the social media posts of a single cancer patient over a 6-month period. The outcome of this study was published in the *BMC Palliative Care* journal (Appendix A.2.2) and presented at the British Computing Society Health Informatics Scotland Conference in the UK (Appendix A.1.4) and at the World Cancer Congress in the USA (Appendix A.1.5).

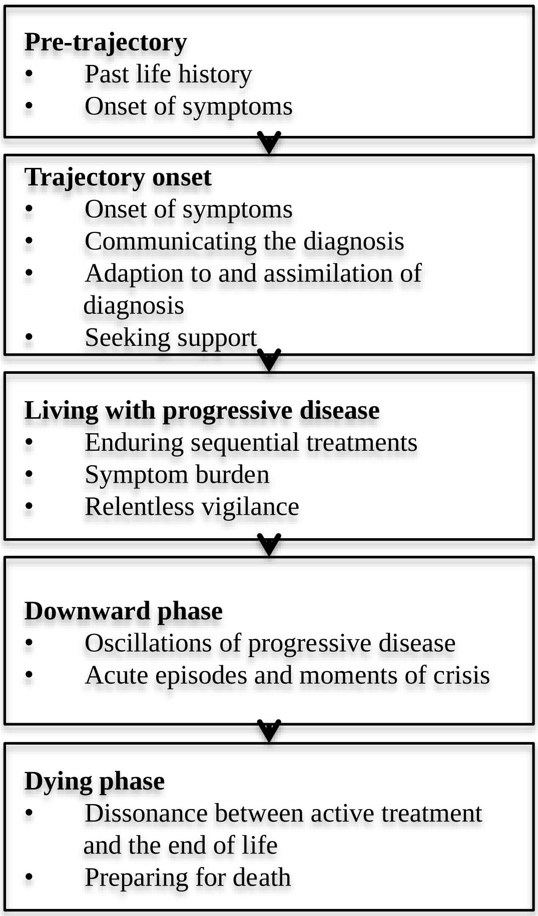
8.1 Context and aim of the study

This study sought to systematically analyse the content of one prominent cancer patient's Twitter feed in the final 6 months of her life, in order to determine its fit with documented end-of-life trajectories and the 6 dimensions of the 'framework for a good death', as well as to explore the value of social media data for understanding patients' personal experiences, life quality and coping strategies. The study was intended as exploratory and hypotheses-generating, with a view to providing insights to inform future research and the design of innovative palliative care services.

In 1968, Glaser and Strauss described the advancement towards death as having the elements of time and shape [460], giving rise to the concept of the illness trajectory. Originally developed to describe how physical aspects of a patient's disease unfold through the phases of pre-trajectory, trajectory onset, living with disease progression, downward phase and dying [461], the concept has since been expanded to include psychosocial aspects of the patient experience, including their response to their illness, the people around them and the interventions undergone by the patient [462]. Emanuel and Emanuel's 'framework for a good death' [439] has been particularly influential in helping clinicians to better anticipate the needs of patients during the progression of their illness and to shape palliative care services. The framework articulates six 'modifiable dimensions' of the patient experience related to 1) physical symptoms, 2) psychological and cognitive symptoms, 3) social relationships and support, 4) economic demands and care-giving needs, 5) hopes and expectations and 6) spiritual and existential beliefs.

Most research into illness trajectories originates from the fields of public health and social sciences, drawing on studies using qualitative or mixed-methods research, with data typically gathered from cohorts of patients through focus groups, surveys and interviews. This research has revealed different illness trajectories for different terminal conditions, with the cancer trajectory described as a steady progression over a period of weeks and sometimes years, punctuated with the positive and negative effects of oncology treatment, weight loss, reduction in physical performance and the impaired ability to self-care during the last few months, as shown in Figure 36 [462, 463]. Cancers can also have unique trajectories, depending on issues such as prognosis, pain, disfigurement and response to treatment [464] while the same type of cancer progression may be experienced differently as a consequence of personal and social factors, such as resilience and availability of emotional support [465].

Figure 36. Reed and Corner's cancer illness trajectory phases [462]



Historically, death and dying has been seen as a taboo topic for public discussion; however, willingness to talk openly about it online by terminally ill patients, their families and palliative health and social care professionals has increased over recent years, giving rise to an emerging field of research [466]. The first empirical study to have analysed data from Twitter for the understanding of disease was published in 2010 [467] and focused on infectious diseases such as influenza and foodborne infections. Among the many subsequent studies using Twitter in such research [468], none of these empirical studies had analysed how the online activity of cancer patients reflects the illness trajectories. Nor had they explored the potential of this emerging data source to yield insights about cancer patients' experiences at the end-of-life.

8.1.1 Kate Granger

Kate Granger was an English geriatrician and campaigner for better patient care. In 2011, at the age of 29, she was diagnosed with a rare form of sarcoma, known as desmoplastic small-round-cell tumour, with a predicted life expectancy of around 5 years. She was treated with P6 protocol chemotherapy and endured painful treatments, which she described in detail in her blogs 'The Other Side and the Bright Side' [469]. Kate created her Twitter account (@GrangerKate) in March 2012 and, prior to her death on 23 July 2016, posted approximately 12,500 tweets and attracted approximately 48,000 followers [470].

Through her experiences as a patient, she and her husband founded the '#hellomynameis' campaign encouraging healthcare staff to introduce themselves to patients. They raised over £250,000 for local cancer charity the Yorkshire Cancer Centre Appeal and, in 2015, Kate was awarded an MBE for her services to the British NHS. As a young woman familiar with social media, Kate's story represents a valuable opportunity to examine the emergence and progression of a personal narrative, in the public domain, about coping with terminal illness.

8.2 Methods

Although the cancer trajectory can last for years and did so for Kate Granger, this study focused on the final 6 months of life, which represents the terminal phase and is commonly associated with preparation for death and the commencement of hospice services, for those fortunate enough to receive the latter [471]. Original tweets, re-tweets and responses posted between 1 January 2016 and 25 July 2016 using the account @GrangerKate were manually extracted for categorisation and analysis. The re-purposed PRISMA guidelines (Chapter 3) were also applied to this empirical study. The data extracted included the date and time of posting and the up to 140-character text contained within the tweet.

The tweets were then manually screened for their relevance to the disease, based on predefined inclusion and exclusion criteria, as described in Box 6.

Box 6. Inclusion and exclusion criteria for Kate Granger's tweets

Inclusion criteria:

- Tweets posted by @GrangerKate's Twitter account
- Original tweets, annotated re-tweets and personal responses posted between 0:00:00 (UTC) 1 January 2016 and 23:59:59 (UTC) 25 July 2016
- Tweets that were considered directly relevant to the terminal condition

Exclusion criteria:

- Tweets posted on other Twitter accounts
- Simple re-tweets of other people's postings (with no further annotation)
- Tweets posted outside the timeframe indicated
- Tweets that were considered not directly relevant to the condition, such as those associated with the '#hellomynameis' campaign, Kate Granger's fundraising activities, her views on the NHS and politics as well as news media
- Images or URLs embedded within tweets

Drawing on principles of digital ethnography [472], qualitative content analysis was used [473] to summarise, chart and interpret the eligible tweets. Tweet content was first categorised according to the six modifiable dimensions of the patient experience in the 'framework for a good death', shown in Table 25, with each post treated as a single unit of interaction and the categories as non-exclusive. The narrative content of tweets was also examined in order to obtain further contextual information about significant events and personal responses, to aid interpretation. The frequency of each category, as well as the occurrence of key events (such as medical procedures or transfer to hospice), were plotted over time and converged with the qualitative data in order to 'learn the meanings, norms, patterns of a way of life' [474: pg13] and to enable comparisons to be made with published end-of-life trajectories. Images and web links included in the tweets were not reviewed during the screening.

Table 25. Modifiable dimensions of the patient experience, from the 'framework for a good death'

Modifiable dimension of patient's experience	**Examples of specific concerns**
Physical symptoms	Pain and fatigue
Psychological and cognitive symptoms	Depression, anxiety and confusion
Social relationships and support	Family, community, interests
Economic demands and care giving needs	Saving and income, personal care and nursing care
Hopes and expectations	Milestones and assessment of prognosis
Spiritual and existential beliefs	Religion, sense of purpose and meaning

The ethical considerations and guidelines described in Chapter 4 were applied to this study. These guidelines included those published by the ESRC, BPS, AoIR and NIHR. The information posted on Twitter is in the public domain and therefore can be analysed for research purposes. However, given that

Kate Granger is the named subject of the study, her anonymity could not be guaranteed and therefore written agreement to extract, analyse and publish the tweets posted by @GrangerKate was sought and obtained from her widower, Chris Pointon, via email.

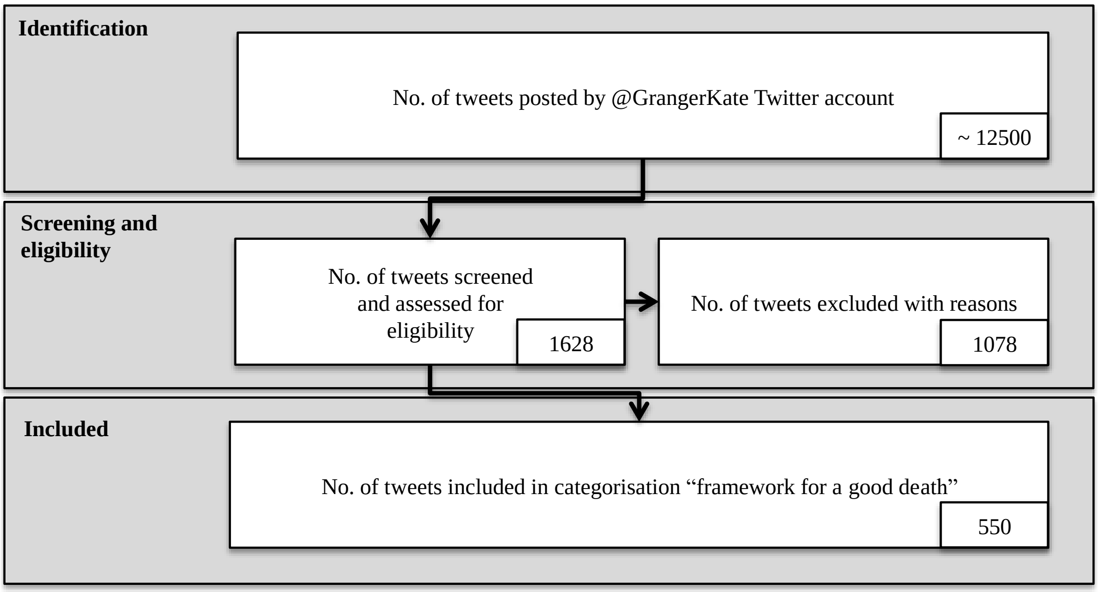
8.3 Results

Kate's tweets were first screened for eligibility and those that were considered relevant to her condition were then classified and plotted on a visual timeline.

8.3.1 Tweet eligibility

Of the 1,628 tweets posted by @GrangerKate during the 6-month period, 550 were considered relevant to her condition and therefore included in the classification, as described in Figure 37.

Figure 37. Steps of screening and assessing Kate Granger's tweets



The remaining 1,078 tweets were excluded for reasons such as their focus on her campaigning activities (n=388, 35.9%) as well as her views on the NHS (n=109, 10.1%) and politics (n=105, 9.7%), as summarised in Table 2.

Table 26. Number of Kate Granger's tweets excluded and the reason for exclusion

Reason for exclusion	**Number of tweets (% of excluded tweets)**
Campaigns including ''hellomynameis' and fundraising activities	388 (35.9%)
Professional work as a geriatrician	89 (8.3%)
Hobbies including baking, flute and band practice	48 (4.5%)
NHS in general	109 (10.1%)
Politics including her support for the Junior Doctors' strike	105 (9.7%)
Dissemination of news and research	38 (3.5%)
Others, including general 'thank you' messages and topics that could not be determined from the tweet narrative	437 (40.5%)

Reasons for exclusion were not mutually exclusive.

The 550 tweets considered relevant to the condition were classified according to the modifiable dimensions of the patient experience in Emanuel and Emanuel's 'framework for a good death'. The total number of tweets classified for each dimension is summarized in Table 27, including an example of the tweet narrative in each category.

Table 27. Number of Kate Granger's tweets per modifiable dimension

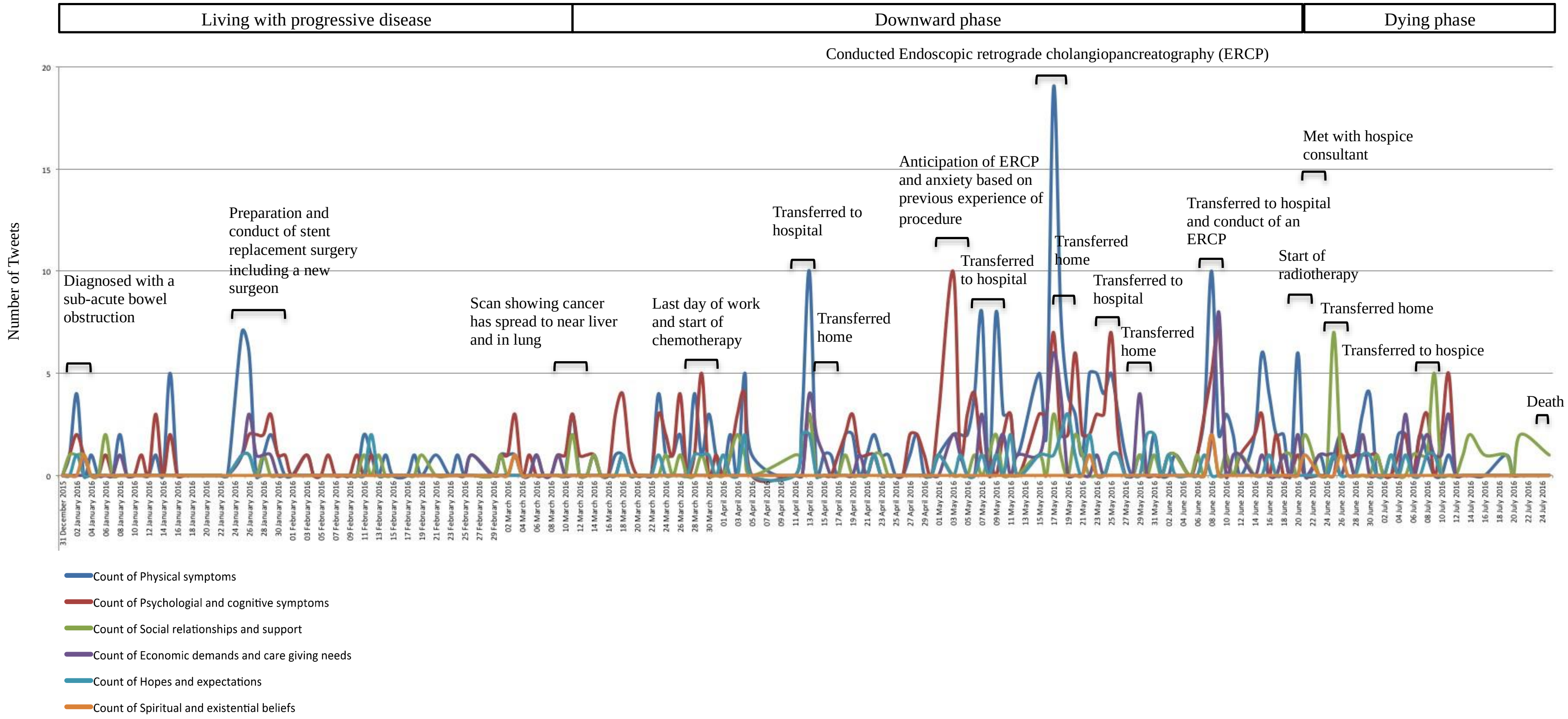
Modifiable dimension of patient's experience	**Number of tweets (%)**	**Examples of tweets**
Physical symptoms	270 (49.1%)	'I am having a temporary nephrostomy this morning... With a view to then getting the radiotherapy done then performing a stent exchange' – 15 June 2016
Psychological and cognitive symptoms	213 (38.7%)	'I'm not sure I can do this.' – 3 April 2016
Social relationships and support	85 (15.5%)	'Everybody being so lovely both in public & behind the scenes on Twitter. Thank you so much. Love our virtual family very much' – 8 May 2016
Economic demands and care giving needs	85 (15.5%)	'After seeing my lovely palliative care nurse this a.m. we've decided hospice admission for symptom control best course of action.' – 8 July 2016
Hopes and expectations	51 (9.3%)	'Perhaps I should just accept #deathbedlive is closer than I hoped it was & get my final preparations finished.' - 29 March 2016
Spiritual and existential beliefs	7 (1.3%)	'A comfortable night, just one would be so appreciated if anyone is listening. Running on empty and a nonstop few days coming up.' – 3 March 2016

Modifiable dimensions of patient experience were not mutually exclusive.

8.3.2 Nature of tweets over time

Figure 38 visualises Kate Granger's digital end-of-life trajectory. The coloured lines represent the frequency of daily tweets, according to each of the six dimensions of the 'framework for a good death', plotted over the 6-month observation period. Similar to Barclay et al.'s study into the trajectories to death in residential care homes [475], the annotations in Figure 38 describe key contextual events and the superscript shows the broad phases of the illness trajectory, which were evident in Kate's tweets.

Figure 38. Kate Granger's illness trajectory



Given that Kate had been diagnosed with cancer for five years, Corbin and Strauss' pre-trajectory and trajectory onset phases were not represented in the sample of tweets. However, tweet characteristics indicative of the living with progressive disease, downward trajectory and dying phases were evident and these form the basis of the findings.

Living with progressive disease phase

The 'living with progressive disease' phase can, in some cases, last several years, where patients go through cycles of feeling and looking well and living life to the full, followed by periods of disease progression and illness. Although the full extent and duration of this phase was not analysed as part of this study, which focused on the final six months of life, initial findings indicate that Kate posted 600 tweets from the beginning of the observation period on 1 January 2016 until 20 March 2016, of which only 16% were associated with her condition. She endured a series of sequential treatments and procedures including her seventh stent replacement in late January 2016, with the symptom burden being indicated by tweets such as '*started 2015 worrying about cancer. Start 2016 worrying about cancer*' and '*we've been living with the reality of cancer in our lives for nearly 5 years…it becomes the new normal*' – 1 January 2016.

Uncertainty seemed to pervade Kate's life during this phase, as indicated by her tweet '*In pain. Horrible achy pelvic pain….Are you waking up cancer or is it in my head?*' – 21 February 2016. She was at home and still able to work, requiring minimal caregiver support, and continued with her hobbies and campaigning activities. She oscillated between disease progression and treatment, with psychological and cognitive symptoms appearing to closely follow the physical symptoms. She had received support from a palliative care team from an early stage ('*Not been to hospice yet but had palliative care team support almost from day one…'* – 4 May 2016). Although treatment allowed some respite from symptoms of progressive disease ('*It kept me well for nearly 6 months. But was never going to cure me & was always a palliative measure…*' – 13 March 2016), living with the condition appeared exhausting in its relentlessness ('*I'm just so tired of being in pain. Nagging pain that I can't escape night after night. Had enough.'* – 21 April 2016).

Downward phase

Kate’s downward phase was characterised by increasingly frequent episodes of illness and crisis over a four-month period between 20 March 2016 and 21 June 2016, which is reflected in a marked increase in the frequency of her tweets. Peaks in this digital trail marked acute episodes, associated with treatments and procedures such as chemotherapy, radiotherapy and endoscopic retrograde cholangiopancreatographies (ERCP) and four stays in hospital over a ten-week period. By this stage of her illness, she was no longer able to work and many of her tweets detailed the physical and psychological anxiety she experienced, as well as the support provided by healthcare professionals and by her physical and virtual community of family, friends and followers. During this phase she posted 898 tweets, of which 42% were associated with her condition.

During these acute episodes, tweets conveyed a lack of confidence in being able to cope (’*I'm exhausted of being 'brave', the expectation that I'm 'Dr Kate Granger' and therefore can cope with anything. I can't...*’ – 3 May 2016) and acknowledging that the end was near (‘*Perhaps I should just accept #deathbedlive is closer than I hoped it was & get my final preparations finished.’* – 19 July 2016). Respite came from being at home (‘*Very happy to be on the way home. Thank you so much to #nhs for scraping me up & putting me back together…’* – 18 May 2016) and the positive messages of thanks and support given by those around her (’*Aww having a little smiley weep at all the wonderful messages. Everyone is just so kind. Thank you so much xx #kateshomecoming*’ – 25 June 2016)

Dying phase

In the four weeks prior to her death on 23 July 2016, Kate experienced a rapid decline in health, which was mirrored by a decline in the number of tweets posted, down to 130, of which 57% were associated with her condition. Kate’s acknowledgement of the terminal phase is indicated in her first reference to the transition to hospice care, with the tweet ’*Hospice waiting list was going to be well into next week so decided to bite the bullet and come home…’* – 23 June 2016. She appeared to be preparing for death during the four-week period after meeting with the hospice consultant on 29 June 2016, before being transferred to St Gemma’s Hospice on 8th July. In contrast to the previous phases, only 30% of Kate’s Tweets in the dying phase referred to her physical and psychological symptoms and the remainder instead focused on the social and caregiving support she received from her palliative care and hospice team and her family, friends and online followers.

8.4 Discussion

This study describes the pattern of physical, psychological, social and care needs of high-profile sarcoma patient Kate Granger, as represented in the posts from her Twitter account over the six-month period prior to her death. The analysis indicates that the digital manifestation of cancer progression maps to the illness trajectories described in existing palliative care research and to the key dimensions of the patient experience articulated in Emanuel and Emanuel's 'framework for a good death'. The analysis has also tested the use of social media data and digital ethnographic approaches for exploring the lived experiences of patients nearing the end-of-life.

8.4.1 Parallels between the digital end-of-life trajectory and existing models

In interpreting the digital trajectory seen in Kate Granger's Twitter postings, this study drew on the general cancer trajectory described in the literature, in which patients experience an onset of incurable disease and a short and rapid decline over a few months. This type of trajectory involves a reasonably predictable decline in physical health, with progression punctuated by the positive and negative effects of palliative oncology treatments [462]. Whilst the analysis focuses on the latter stages of the cancer illness trajectory, it is important to consider the similarities and differences between the patterns seen in Kate Granger's tweets and these general trajectory descriptions. In Kate's case, similarities can be drawn in relation to her endurance of sequential treatments and the acute episodes and moments of crisis, as well as her preparations for death. As noted earlier in this chapter, the trajectory of different cancers varies, to some extent, and it is important to bear this in mind when seeking to interpret the patterns observed in individual cases. For example, Reed and Corner's research into the illness trajectory of metastatic breast cancer identified a 'rollercoaster' trajectory, where the typical duration was 2 to 5 years and patients, similarly to Kate Granger, experience oscillations of disease progression, treatment and restoration of well-being, as well as acute crisis episodes [463].

Emanuel and Emanuel's 'framework for a good death' proved a useful tool for classifying the data, which showed a good fit with their six 'modifiable dimensions of the patient experience' and helped in the plotting of physical symptoms and psychosocial responses. For example, the differentiation of tweet types aided the interpretation of changes in Twitter activity between the different phases of the illness trajectory, such as the priority accorded to physical and medical issues in the earlier stages, compared to the focus on people and comfort-giving at the end.

It is interesting to note that, based on her tweets over the 6-month period, Kate made no explicit reference to spirituality, which contrasts with previous observations in palliative care research [476]. Despite this, there were frequent references to metaphysical concepts, such as the call for ‘someone out there’, suggesting that existential concerns may simply be taking new forms with the decline of traditional religious practices in UK society.

Also absent in Kate’s Twitter narrative are ‘battle metaphors’, such as talk of fighting or being at war with cancer. While these are common in some cultural and clinical settings and have been the subject of research [477], experts have advised healthcare professionals against using them with patients, to avoid inducing feelings of failure for what is a biologically-determined outcome [478]. As a health professional, Kate Granger would have been acutely aware of her prognosis and may thus have chosen to focus on coping and preserving her quality of life. This is also consistent with the importance of recognising the important influence of individual patient characteristics in how a ‘good death’ is experienced [439].

8.4.2 Benefits of analysing social media data

This study takes a first step in demonstrating how these emerging data sources may elucidate terminal patients’ physical and psychosocial responses during the illness trajectory and thus help to inform the provision of supportive and palliative care services at different stages. Analysing the social media trail of an individual like Kate Granger provides a unique window into the ‘lived experience’ of real people in real time, including at highly emotional and sensitive stages, which can be difficult to access using conventional direct research methods. Social media are becoming the norm for communication amongst younger people but, as these ‘digital natives’ progress to later stages of life, where death becomes more imminent, there are likely to be greater opportunities to study larger cohorts of patients with different terminal conditions. Social media data are useful not only for focused studies, such as the study undertaken here, but also present opportunities to undertake research at scale. In the only other study that was found to have charted the cancer trajectory for sarcoma, Tang and colleagues converged electronic record linkage with individually-administered questionnaires, in order to profile levels of distress in as many as 74 patients before, during and after surgery, with the aim of understanding the psychological and socioeconomic factors influencing resilience, coping and outcomes [464]. The rapid convergence of eHealth research, data science and digital social research offers the potential for social media mining and analytics to enable both types of research, as well as to understand broader patterns of the death experience in society. The use of automated analysis and natural language processing tools are widely available and are used to monitor brand, reputation, content and sentiment on social media [479]. They include those provided by Crimson Hexagon, Hootsuite, Symplur, Keyhole and Sproutsocial. Although predominately applied in the context of marketing activities, they could be re-purposed for use in the kind of research described in this study.

8.4.3 Implications for supportive and palliative care

Illness trajectories have proven valuable as a means of elucidating the psycho-physical progression of cancer and other conditions [462-464]. While mapping the objective physical and medical aspects of these journeys is relatively straightforward, mapping their psychological and socio-emotional aspects has required lengthy in-depth qualitative research, primarily using interviews with patients and their loved ones [462], which has constrained the usefulness of the insights for practitioners. Being able to study the journey towards death in the digital world opens a new window into the concerns, needs and vulnerabilities which dying patients experience at different points in time and can help to target the provision of supportive and palliative care, as well as help health professionals to understand patients' perspectives on the care they deliver. Existing research has shown how insights from different trajectories may inform each another [462]. Combining new approaches to public health surveillance and automated analysis of data from social media sites may also generate new insights about the common and unique patterns characterising the progression of different life-limiting illnesses.

The longitudinal data posted by Kate Granger also provide evidence of the involvement of palliative care services during the patient journey. In Kate's case, she received palliative care early on in her diagnosis with her first tweet making reference to this being posted on 25 October 2012, 14 months after her initial diagnosis. This is later supported by her tweet *'…had palliative care team support almost from day one. They've been amazing... See the whole of me*' – 4 May 2016. While palliative care services still tend to focus on the shorter dying phase [463], research has demonstrated that engagement over longer trajectories can enable better advanced planning for a good death, empower patients attempting to gain control over their illness and help to alleviate concerns about the possible nature of death [462]. All of these were evident in Kate's tweet '*After seeing my lovely palliative care nurse this a.m we've decided hospice admission for symptom control best course of action...*' - 8 July 2016.

End-of-life care planning must be multi-dimensional, with palliative care services playing the role of a mediator in helping patients to cope with their illness, optimising quality of life and achieving a dignified and peaceful death. Previous research, including that focused on sarcoma patients, has revealed a range of problem-focused coping strategies, such as information seeking, choosing one's treatment team and advocacy for oneself, as well as emotion-focused strategies such as support seeking, present-moment focus, distraction, denial and oversleeping [465]. Future analyses of patients' social media activity may help to verify or shed further light on these strategies, in addition to profiling illness trajectories and dimensions of the 'good death' framework. It may also help practitioners to better understand differences in patients' responses to their illness. For example, based on the analysis of Kate Granger's data, it appears that she did not use Twitter as a problem-focused coping strategy as there were few questions in her tweets suggesting a lack of information or opinion seeking interactions with her followers. This could be explained by the focus on the last 6 months of her life, rather than the period of initial diagnosis, or by her professional role as a geriatrician, which put her in a more informed role than most other patients. Understanding patients' trajectories of need can also help palliative care professionals to better anticipate and proactively mitigate distress [476]. In Kate's case this need for responsive approaches was evident in tweets such as '*Amazing care from #NHS today with my port flush, blood tests & psychology appointment…*' - 26 February 2016 and '*Psychology appt could not have come at a better time*' – 18 March 2016.

Another way in which digital ethnographic research methods may complement existing illness trajectory research is through their potential to reveal the additional support provided to patients by their online social networks and communities. Kate Granger had many online followers, who helped to lift her spirits during periods of difficult treatment and distress, and studying the patterns of support and reciprocity in these digital spaces may help to suggests new ways in which to help patients nearing death. Despite Kate's predicament, she did not appear to be lonely, as can often be the case, and regularly thanked her followers for their support during acute episodes of crisis. As such, palliative care teams may consider recommending that terminal patients establish an online presence on social media and share their experiences with others, as an emotion-focused coping strategy and means of providing social and psychological support in addition to that received more directly from family, friends and care services. Online spaces can however be unsafe for users and therefore patients should also be informed of the potential dangers that exist, which social media platforms to use and how to operate within them.

New approaches to using social media also offer opportunities to study patients' trajectories even after the end-of-life. The ability of individuals to preserve their digital legacy [480] is becoming increasingly common, with various services now available, such as DeadSocial, LifeNaut and Afterword, to enable users to retain on a perpetual online version of themselves after death, in the form of a social media auto-biography or even as an avatar that can be accessed and can interact with their friends and relatives [481]. This opens new sociological and ethical questions about the responsible management of personas after death and about the donation of social media archives for future research purposes, in a similar way to the donation of medical records, tissue samples or body parts [482].

Limitations

This cohort size of one limits the extent to which the findings of this study can be generalised to other patient groups. The implications of this are that the findings from this study may be unique to Kate Granger. Further comparable research is recommended in order to extend the evidence-base, including studies exploring whether the end-of-life trajectories reported for different people with the same condition as well as those with different conditions are also evident on Twitter or other social media platforms [476].

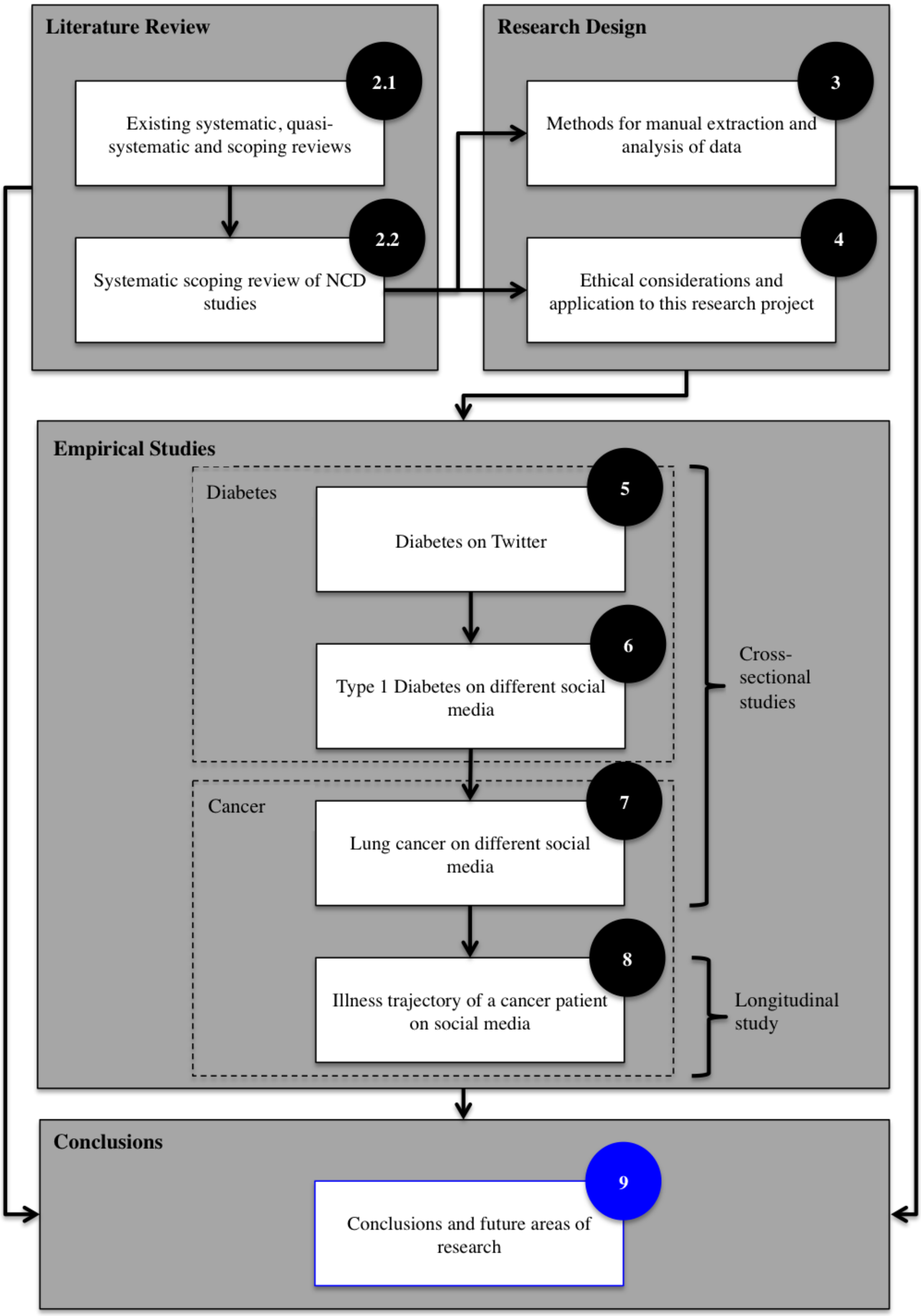
The entire sample of 12,500 tweets from the four years that the @GrangerKate Twitter account was active was not extracted and analysed during this study, limiting the conclusions drawn with regards the complete illness trajectory from diagnosis to death. This could be addressed through sampling for 3-4 months across each year of the illness and undertaking a similar analysis to this or the use of automated social media mining techniques and natural language processing tools. Nevertheless the use of automated techniques also present challenges, including variability in the types of social media to which they are suited and the scope of data available for analysis, the cost of software and data access licences and the transparency and accuracy of algorithms used to generate inferences from social media data.

Despite numerous studies into what constitutes a good death, there is little agreement about its definition. Key features have been identified as: preferences for a specific dying process, pain-free status, religiosity/spiritualty, emotional well-being, life completion, treatment preferences, dignity, family, quality of life and relationships with healthcare providers [483]. Although valuable and well respected, the 'framework for a good death' goes only some way towards accounting for the complexity of the end-of-life experience. In this study, the different dimensions were linked and therefore, in some cases, were difficult to distinguish as part of the analysis.

The data on social media sites also presents challenges, due to the self selected nature of contributors such as Kate Granger as a Twitter account is required to post content, the fact that not all patients are able to or willing to share their experiences online, the justifiable anonymity preferred by many users and the uncertain provenance and accuracy of the information posted.

8.5 Conclusions

This is believed to be the first known study to have systematically analysed the end-of-life illness trajectory expressed in a patient's social media activity. The results indicate that the data posted by terminal patients on Twitter and other social media can provide insights that may be comparable to or complement those garnered using more traditional qualitative research techniques. To quote Kate, '*the power of patient narrative cannot be underestimated*'. While the analysis was at the structured end of the digital ethnographic spectrum, it nevertheless shows the value of such methods for understanding how terminal disease is experienced by and affects individuals, how they cope, how support is sought and obtained and how patients feel about the ability of palliative care services to meet their needs at different stages. More research is warranted to extend this analysis across the wider trajectory of life-limiting illness and to variety of disease types, as well as to explore the use of informatics-driven data mining and pattern recognition techniques to study larger cohorts and different social media types, including social networks like Facebook. As part of a wider agenda for 'palliative social media' [466], it is recommended that efforts be made to engage health professionals in exploring how digital end-of-life trajectories may inform the provision of supportive and palliative care, to improve the quality of life and death for patients like Kate.


Literature Review
2.1
Existing systematic, quasi-systematic and scoping reviews
2.2
Systematic scoping review of NCD studies
Research Design
3
Methods for manual extraction and analysis of data
4
Ethical considerations and application to this research project
Empirical Studies
Diabetes
5
Diabetes on Twitter
6
Type 1 Diabetes on different social media
Cancer
7
Lung cancer on different social media
8
Illness trajectory of a cancer patient on social media
Cross-sectional studies
Longitudinal study
Conclusions
9
Conclusions and future areas of research

9. Conclusions and areas of future research

The aim of this PhD research project was to address the gaps in empirical, evidence-based research that source data from social media to understand patient health issues and inform public health research into NCDs. Seven research questions were defined and investigated to systematically scope the status of existing multi-disciplinary research literature (Section 2.2, RQ1), inform the manual methods for extracting and analysing data from social media (Chapter 3, RQ2) and address the ethical considerations of conducting such research (Chapter 4, RQ3). The four empirical studies (Chapter 5-8) that were conducted contribute to a growing body of knowledge into the dynamics and value of patient narratives on different social media for different NCDs.

The systematic scoping review of the literature (Section 2.2) highlighted a rapid increase in the number of studies published in the last five years, with 87% of total empirical studies published between 2013 and 2017. This poses a challenge to researchers conducting studies in this fast-moving and evolving field, where regular and timely publication and presentation of findings is needed to mitigate the risk of duplicative and redundant research. Given the volume and speed at which empirical studies are published, it is pertinent that automated methods for conducting systematic and scoping reviews [258, 259] be considered. This will enable the staggering number of empirical studies to be analysed in a pragmatic manner and at regular intervals over the coming years, in order to identify gaps and inform the direction of future research that source data from social media. A comparison between the results of automated and manual systematic review methods is recommended to inform their accuracy.

The number of NCD-related studies has also increased over this period, albeit at a slower rate with only 15% of the 695 studies that source data from social media focus on NCDs and that of the 105 eligible studies, cancer and diabetes were investigated in 54% and 20%, respectively. This finding aligns with previous systematic, quasi-systematic and scoping reviews discussed in Section 2.1, which highlight the lack of empirical studies, as well as the importance of conducting and repeating reviews at regular intervals in order to draw conclusions about the evolution of the field.

Type 1 diabetes and lung cancer are underrepresented in the existing literature (Section 2.2), a gap into which findings from the empirical studies described in Chapters 6 and 7 provide insight. Users of social media may not necessarily represent the population of people living with NCDs, given the ‘digital health divide’ [394] and the higher prevalence of diabetes [395] and lung cancer [454] in older age groups. Although research exists into this divide and the consequences for research that source data from social media [252-254], little is known about its implications for studies that investigate NCDs. A deficit of research into cardiovascular and respiratory conditions, other chronic conditions and certain types of cancer (e.g. kidney, liver, skin, stomach) continues to exist, presenting opportunity for further novel research studies.

Additional findings from the systematic scoping review highlight that most of the studies (92%), including those described in this thesis, sourced English-language posts or posts from countries that are predominantly English speaking and were conducted in high-income countries (97%). The limited number of NCD studies that investigate population segments with different demographic cohorts and geographical locations points to a need to further understand the implications of the ‘digital health divide’. This is in line with the WHO’s *draft global strategy on digital health 2020-2024* [4], and the inclusion of studies published in languages other than English, to provide further insight into the social media narrative. This insight can be used to inform policy and public health interventions as well as access and standards of health services.

The increasing use of automated methods correlates with the increasing number of studies that extract and analyse large samples of data from social media, highlighting how the field is evolving. The application of computer science and informatics capabilities points to a convergence of disciplines that researchers, institutions and funding bodies must navigate to provide timely and relevant insights and guidance. This convergence of academic disciplines was reflected in this thesis and associated studies, by the co-supervision of this PhD research project by faculty members from the inter-disciplinary research community at the Usher Institute of Population Health Sciences and Health Informatics and the eHealth research group, that sit within the College of Medicine and Veterinary Medicine at the University of Edinburgh as well as the Informatics department.

The studies described in this thesis focused on text content posted on social media and do not provide insight into the visual or audio content shared. When this PhD research project was initiated in 2013, manual techniques were used in 42% of the eligible NCD studies without consistent guidance or application of methods. The experience gained from re-purposing and applying the PRISMA guidelines and checklist to the manual analysis of social media data was described in Chapter 3 (RQ2). It was concluded that, although the qualitative data posted on social media is primary data, many of the requirements for a systematic review could be applied with the acknowledgement of limitations related to the assessment of bias. These insights were presented at the Medicine 2.0 Summit and World Congress in 2014 as a means of informing other researchers conducting similar studies.

Findings from the systematic scoping review of literature revealed that the frequency of automated methods and tools for extracting and analysing data from social media is increasing. There is, however limited insight into the accuracy of their application in the context of NCDs or when sourcing data from different social media platforms. Further research is warranted that compares the results from extracting and analysing identical data samples when applying manual approaches, automated methods or a hybrid, the results of which can be used to better inform the research design of future empirical studies with increasingly larger samples of data.

RQ3 (Chapter 4) exposed the deficit of ethical guidance for UK based researchers conducting studies that extract and analyse data from social media. This seminal research was published in *Research Ethics* in 2017 [484] and discussed in The Times and other news outlets (Appendix A.2.2.1). Although useful for informing researchers on designing and conducting ethical empirical studies, the review of ethics guidelines was limited to those provided or recommended by RCUK and its seven UK Research Councils. This included the guidelines published by the ESRC, BPS, AoIR and NIHR. Despite their application in this PhD research project, further research is warranted that involves a wider corpus of guidelines such as those published by funding bodies in other countries and research institutions where this type of research is being conducted, as well as journals and conferences where the research is published. New empirical studies are also needed to shed light on the ways in which such communities are addressing approval requests for studies involving the reuse of data from social media. The increasing use of automated methods has since 2017, led to the introduction of terms such ‘ethical NLP’ [485] and ‘ethical AI’ [486], which require further research and guidance to inform communities that apply these methodological innovations.

The two empirical studies investigating diabetes to address RQ4 and RQ5, (Chapters 5 and 6, respectively) provided insights into the narrative associated with different types of diabetes on Twitter as well as the affordances and utility of different social media for those living with Type 1 diabetes. The systematic scoping review of literature (Chapter 2.2) identified Twitter (26%) and Facebook (14%) as the most popular sources of data for eligible NCD studies. Twitter, however, is a ‘noisy’ source of data and although it, and Facebook and Diabetes.co.uk, are all used to disseminate information, Facebook and Diabetes.co.uk are also used for social interaction and peer support. The results of these studies were presented at the Stanford Medicine X conference in 2014 and published in the conference proceedings of the International Conference on Global Health Challenges in 2015 [434].

Since 2013, one study [163] was identified during the systematic scoping review that investigated diabetes using a longitudinal study design. Given the growing number of people diagnosed with diabetes [487], further research is necessary into disease surveillance for NCDs to better inform interventions affecting prevention, diagnosis and management of the condition, as well as the behavioural change and support required by those living with it. Longitudinal disease surveillance of populations requires the extraction and analysis of large data sets on a real-time and continuous basis using automated methods and therefore further empirical research is needed to build on the findings from the existing studies.

To investigate RQ6, the comparison of Twitter, Facebook and discussion forum on Macmillan.org.uk by those affected by lung cancer (Chapter 7) provides new insight into the social and supportive value of different social media, as well as how interactions may be shaped both by the configuration and moderation of the platforms and by users self-organising around groups or hashtags. It highlights the importance of selecting the appropriate social media for potential interventions and to address specific research questions as well as in providing empirical evidence that people affected by lung cancer should consider online communities as an additional source of social support during times of crisis. The terms used to describe the different forms of support offer potential for more in-depth qualitative analysis and to create greater precision of automated methods for extracting and analysing data from social media, which warrant further research. The uniqueness of this study led to its publication in *npj Digital Medicine* in 2019 [488] and underlines the scarcity of studies into lung cancer narratives on social media that was evident in the results of the systematic scoping review (3%). None of the studies to date were longitudinal in design or considered the different illness trajectories associated with this condition. This presents a deficit of literature creating further opportunities for novel future research that are planned after the completion of this PhD research project.

By investigating whether illness trajectories are reflected in a terminally ill cancer patient’s digital narrative to address RQ7 (Chapter 8), this revealed that patient activity on social media mirrors the different phases of the end-of-life illness trajectory described in the literature and that it is comparable to or complements insights garnered using more traditional qualitative research techniques. It also shows the value of such innovative methods for understanding how terminal disease is experienced by and affects individuals, how they cope, how support is sought and obtained and how patients feel about the ability of palliative care services to meet their needs at different stages. This seminal study was published in *BMC Palliative Care* [447] and presented at the 2017 British Computer Society Health Informatics Conference and 2018 World Cancer Congress. Further research is necessary to extend this analysis across a wider trajectory of life-limiting illness and disease types as well larger study cohorts and different social media types. Such a study is planned after the completion of this PhD research project, with the goal of understanding how the end-of-life experiences of multiple lung cancer patients are represented in the narrative on Twitter and their personal blogs. Additional empirical studies that source data from social media are also warranted to build on the growing body of research into survivorship [489] and grief [490, 491].

In conclusion, the secondary use and analysis of health data posted on social media continues to be an expanding and evolving field of multi-disciplinary research. The studies described in this thesis provided timely guidance to researchers who use manual methods for extracting and analysing this data and recommendations for applying existing ethics guidelines. The seminal empirical studies revealed insights into the utility and affordances offered by different types of social media and the forms of support sought and received by those affected by diabetes and cancer, hence demonstrating the value of social media narratives to patients, communities and health researchers. This thesis and the associated publications contribute to an exciting and emerging body of knowledge around NCDs and eHealth that is informing public health practice, interventions and policy, as well as offering direction for further innovative and novel studies to inform the field.

Appendices

A.1 Conference abstracts, posters and papers

The PhD research project described in this thesis was conducted over a six-year period. The research studies described in Chapters 3, 5, 6 and 8 addressing RQs 2, 4, 5, and 7, respectively, was presented at conferences and published in conference proceedings.

A.1.1 Stanford Medicine X Conference, USA (2014)

Date: 5-7 September 2014

Location: Stanford, California, USA

Type of presentation: Research-in-progress presentation

A.1.1.1 Abstract submitted to conference

Title: An exploratory study into how the clinical condition of diabetes is discussed on Twitter.

Background: Technological advances over the past decade have resulted in increasing numbers of the global population having access to the Internet and, in particular, social media sites such as Facebook, Twitter and YouTube, as well as health- specific social media sites such as PatientsLikeMe. Although not solely related to public health, the increasing use of the Internet to search for health-related information and connect and communicate with patients and healthcare professionals, as well as express opinions and share experiences, has lead to an influx of public health research into its potential use.

Twitter, in particular, has been used as a source of data in numerous public health research studies aimed at understanding the online support, behaviour and information patients receive, as well as for monitoring disease outbreaks. Research studies published by the Pew Research Center indicate that 18% of online adults use Twitter compared to other social media sites such as Facebook, Instagram, Pinterest and LinkedIn and that diabetes was the condition ninth most commonly searched in 2010 by people looking for information online. Despite this, there are few studies that have explored how the topic of diabetes is discussed on social media sites such as Twitter, by whom and for what purpose.

Objective: To assess the relevance of diabetes-related postings on Twitter to the clinical condition of diabetes and the topics and intentions represented in these messages.

Methods: This exploratory study applies the PRISMA guidelines for conducting systematic reviews to the identification, screening and selection of diabetes-related Tweets that were posted in September 2013. This data was extracted from Twitter and analysed in a two-stage process. First, the search term ‘diabetes’ was used and tweets were screened and selected based on predefined inclusion criteria. Recurrent topics were identified using the constant comparison method across four categories; characteristic of the tweet, content of the tweet, intention of the tweet and stakeholder group for which the tweet is relevant. This will be analysed using content analysis to classify and count the tweets, as well as cluster analysis techniques in order to represent their statistical associations and differentiations.

This is research in progress and the results and analysis are currently under way.

Presentation is available on YouTube: https://www.youtube.com/watch?v=UCrCWcWQmQk

A.1.2 Medicine 2.0 Summit and World Congress, USA (2014)

Date: 13-14 November 2014

Location: Maui, Hawaii, USA

Type of presentation: Poster

A.1.2.1 Abstract submitted to conference

Title: Applying Systematic Review Methodologies to the Analysis of Data Available on Social Media Sites: An Exploratory Study into How the Topic of Gestational Diabetes is Discussed on Twitter

Background: The use of social media for health-related activities is widely accepted to be on the increase. There is limited guidance, however, on how to systematically review the data available on social media sites in the context of health. Although systematic reviews are becoming increasing important in public health research, there are currently no known examples of its application to the data available on social media sites.

Objective: To evaluate whether the widely applied PRISMA guidelines for conducting systematic reviews can be repurposed and applied to the data available on Twitter. This will be achieved through conducting an exploratory study to assess the relevance of gestational diabetes-related postings to the clinical condition of gestational diabetes and the topics and intensions represented in these messages.

Methods: This paper applies the PRISMA guidelines for conducting systematic reviews to the identification, screening, selection and synthesis of diabetes-related Tweets that were posted in September 2013. This data was extracted between May and November 2013 from a 1% Twitter sample. The sample was analysed in a three-stage process. First, the search term ‘diabetes’ was used and tweets were screened and selected based on predefined inclusion criteria. Recurrent topics were then identified using mixed methods of synthesis. Finally, these themes were then applied to the gestational diabetes-specific tweets. Based on the experience gained from conducting the exploratory study into gestational diabetes, the PRISMA guidelines were later reviewed to determine their applicability and limitations when applied to the data available on social media sites.

Results: The first stage of the process identified 2,888 Tweets that referred to diabetes. These data were screened for their relevance to health and 1,870 were selected for further analyses. This, in turn, identified 23 conceptual codes across 4 categories; characteristics of the tweet, content of the tweet, stakeholder group for which the tweet is relevant and intention of the tweet. The final stage indicated that gestational diabetes was only referred to in 31 tweets and these only related to 12 of the identified themes.

Conclusions: Given the limited number of conceptual codes that were present in the data sample, initial findings from the exploratory study suggest that Twitter is not the social media site of choice for authors to post or consume content in relation to the clinical condition of gestational diabetes. The experience gained from applying the PRISMA guidelines concludes that, although the data available on social media sites is primary data, rather than secondary data in traditional systematic reviews, many of the requirements can be repurposed and applied to qualitative health data available on social media sites. It is hoped that this study will provide initial guidance on how to systematically review the content of social media sites, as well as facilitate further discussion and research into how social media sites are used to discuss different health-related topics.

A.1.2.2 Poster

Applying Systematic Review Methodologies to the Analysis of Data Available on Social Media Sites: An Exploratory Study into How the Topic of Gestational Diabetes is Discussed on Twitter

Joanna Taylor*+, Miles Osborne**, Claudia Pagliari*

*Center for Population Health Sciences, University of Edinburgh, UK
**Department of Informatics, University of Edinburgh, UK
+EY

Introduction

Increasing use of the Internet to search for health related information has lead to an influx of public health research into its potential use

Twitter has been used as a source of data in numerous public health research studies aimed at understanding the online support, behaviour and information patients receive in relation to cancer (1-3), dementia (4), epilepsy (5,6) as well as in the dissemination of public health information related to vaccines (7), sexual health (8), alcohol consumption (9, 10) and smoking (11, 12). Twitter data has also been used extensively to monitor disease outbreaks (13-17).

There is limited guidance however on how to analyze the qualitative data available on social media sites.

Objective

To evaluate whether the widely applied PRISMA guidelines (18) for conducting systematic reviews, can be repurposed and applied to the data available on social media sites.

Conclusions

- Data available on social media sites is primary data as opposed to the secondary data used in traditional systematic review
- Many of the requirements for a systematic review of literature can be re-purposed and applied to qualitative health data available on social media sites
- This study provides initial guidance on how to systematically review the qualitative data on social media sites
- Further research required to compare its effectiveness with other techniques and methods for reviewing and analysing data from social media sites

Re-purposing PRISMA

Applicability and Limitations

Section/Topic	Item No.	Applicability and limitations when used to review data on social media sites based on the experiences gained from conducting the exploratory study
Title		
Title	1	Applicable
Abstract		
Structured summary	2	Applicable
Introduction		
Rationale	3	Applicable
Objectives	4	Applicable
Methods		
Protocol and registration	5	Applicable
Eligibility criteria	6	Applicable
Information sources	7	Applicable although the exploratory study into how the topic of gestational diabetes is discussed on Twitter, was limited to only one social media site and could have been extended to include multiple social media sites
Search	8	Easily applicable to Twitter data. There may be limitations when considered in the context of other social media sites and the accessibility of the data.
Study selection	9	Applicable to the exploratory study described in this paper, due to the search criteria, pre-defined inclusion and exclusion criteria and character limit of tweets. It could, in this example, be considered more simplistic when applied to social media data compared to the studies traditionally included in systematic reviews. This, however, would need to be reconsidered in the context of other social media sites such as blogs and discussion forums.
Data collection process	10	Not applicable to social media as already described in item number 8
Data items	11	Applicable
Risk of bias in individual studies	12	The risk of bias in individual social media data points is dependent on the type of social media data included in the review. The data from Twitter is limited by character length and therefore risk of bias within individual data points in this case is limited.
Summary of measures	13	Was not applied in the exploratory study and therefore this limits the conclusions that be drawn as to its applicability
Synthesis of results	14	Applicable although it would be dependent on the type of social media data under review (e.g. blog entries, videos) as to which method of synthesis is most relevant.
Risk of bias across studies	15	Applicable
Additional analyses	16	Applicable
Results		
Study selection	17	Applicable
Study characteristics	18	Applicable
Risk of bias within studies	19	The risk of bias within the social media data points is dependent on the type of social media data included in the review.
Results of individual studies	20	Applicable to the classification of individual data points
Synthesis of results	21	Dependent on the type of social media data under review. A formal meta-analysis was not conducted for the exploratory study into how gestational diabetes is discussed on Twitter and therefore this limits the conclusions that can be drawn as to its applicability.
Risk of bias across studies	22	Risk of bias across social media data points could be applied through a risk assessment of the social media data. This however was not considered as part of the exploratory study described in this paper, except to understand when an organization or community authored the tweet.
Additional analysis	23	Applicable
Discussion		
Summary of evidence	24	Applicable
Limitations	25	Applicable
Conclusions	26	Applicable
Funding		
Funding	27	Applicable

Gestational Diabetes on Twitter

Summary of the data

- Random 1% sample of available tweets extracted on 26 October 2013
- Duplicates and spam were removed
- Inclusion criteria:
  - English
  - Contains the word "diabetes"
  - Posted between 0:00:00 (UTC) on 1 September 2013 and 23:59:59 (UTC) on 30 September 2013
  - Classified as health related

Qualitative Analysis

- 24 codes identified through iterative refinement of the classification structure (21 of which are relevant to Gestational Diabetes)
- Codes are not mutually exclusive

Results and Findings

- Majority of tweets within the sample referred to the diagnosis, treatment or management of the condition
- Primary reason gestational diabetes is discussed on Twitter is as a means of sharing information
- There are few references to companies or brands and no reference to any of the medications used to treat the condition
- Twitter is not necessarily the optimal social media site for gaining patient insight in relation to the clinical condition of gestational diabetes

References

A.1.3 International Conference on Global Health Challenges, France (2015)

Date: July 19 - 24, 2015

Location: Nice, France

Type of presentation: Research presentation

A.1.3.1 Short paper published in conference proceedings

Understanding the social dynamics of Twitter, Facebook and Diabetes.co.uk and their value implications for patients and health researchers

Joanna Taylor

Centre for Population Health Sciences,
University of Edinburgh, United Kingdom
Ernst and Young AG, Switzerland
email:
joanna.taylor@ed.ac.uk

Claudia Pagliari

Centre for Population Health Sciences,
University of Edinburgh, United Kingdom
email:
claudia.pagliari@ed.ac.uk

Miles Osborne

Bloomberg LP,
United Kingdom
email:
mosborne29@bloomberg.net

***Abstract*—Health and wellness are dominant societal concerns, which is reflected in their presence on Social Media. People with diabetes use a range of Social Media to share information, build knowledge and seek peer support, but surprisingly little is known about how this behaviour varies across platforms. We drew data from a 10 day period in September 2014 from Twitter, Facebook and the Diabetes.co.uk discussion forum and classified these according to their informational and social properties, using Bales Interaction Process Analysis (IPA). Contrary to the generalisations made in previous research, Twitter is chiefly used for information dissemination, whilst Facebook and Diabetes.co.uk are also used for social interaction and peer support. These differences exist due to the structure of these platforms, including the potential for threaded discussions, the specificity of the user base and the presence of a moderator, each of which influence the nature of member interactions. Our novel findings contribute new insight about the social function of different Social Media in healthcare and the relative value of these Social Media as sources of data for health research, tools for health promotion and intervention, as well as forums for community and patient engagement.**

Keywords–health, diabetes, social media, social network, Facebook, Twitter

I. INTRODUCTION

Type 1 diabetes is a chronic autoimmune condition, which occurs when the pancreas stops producing insulin. This results in increased levels of glucose in the blood, putting patients at long-term risk of heart disease, stroke, nerve damage, kidney disease and blindness. There is currently no known cure for Type 1 diabetes and those diagnosed are treated either by insulin injections, insulin pump therapy, islet cell transplantation or a pancreas transplant.

Hamm et al. (2013) concluded that patients most commonly use Social Media as a means of supporting selfcare and that the literature is dominated by studies of online discussion forums and support groups, followed by dedicated social networking sites and micro-blogs [1]. Social Media has become an increasingly popular data source for public health researchers to understand how members of patient communities interact with each other regarding specific conditions [2], [3].

Previous research has focused on single platforms such as Twitter, Facebook or condition specific online communities [3], [4], [5]. We go beyond such studies with a view to gaining insights into the social interactions that occur across and within Social Media. We are not aware of any previously published study to have compared Twitter (twitter.com) Facebook (https://www.facebook.com/T1Diabetes) and Diabetes.co.uk (http://www.diabetes.co.uk/forum/category/type-1-diabetes.19/) in relation to Type 1 diabetes. Motivated by this, we considered the following research questions, the answers to which will help inform the design of future successful Social Media for the purposes of health and will help patients understand better how they can manage their conditions using them:

RQ1: Do diabetes Type 1 patients use different Social Media platforms for different purposes?

RQ2: Which Social Media are successful at encouraging social interaction and support for patients?

RQ3: What are the implications for Social Media design?

Our three Social Media embody different styles of social interaction. Twitter and Facebook are well known, general purpose sites. Diabetes.co.uk is a condition specific discussion forum where users can create content and others can comment.

The rest of this paper is organised as follows. Section II describes the methodology and datasets. Section III describes outcome of RQ1. Section IV describes the outcome of RQ2. Section V describes the outcome of RQ3. Section VI provides a conclusion to close the article.

II. METHODOLOGY AND DATASETS

For our comparison of Social Media use by diabetes patients, we took a two-stage approach: firstly we extracted and screened posts made by 533 users from each of the three sites and then categorised them using the Bales IPA categories.

ISBN: 978-1-61208-424-4

Application of Bales Interaction Process Analysis Bales IPA system [6] was first introduced in 1951 and has been widely used in public health research as a means of identifying and recording the nature, not the content, of group interactions. Bales identified and defined twelve categories of group interaction each of which were considered when reviewing the Type 1 diabetes posts. Each post was considered a single unit of interaction and that the categories were not considered mutually exclusive when applied to the sample of posts.

Twitter A random 1% sample of all available Tweets was extracted on the 3 October 2014. The tweets were posted between 0:00:00 (UTC) on 1 September 2014 and 23:59:59 (UTC) on 10 September 2014 and were extracted by crawling the data through the standard publically available Twitter API using the pre-defined search terms *diabetes, type 1 diabetes, t1 diabetes, t1d* and *type 1*.

The data extracted included the date and time of posting in UTC, the Twitter account id and the text in the tweet. Re-tweets were identified and any duplicates and spam were removed. The sample of 1433 English language tweets was manually screened. Those that referred to Type 1 diabetes, produced a sample of 66 posts, which were subsequently included in the categorization stage. url links included in the tweets were not reviewed during the screening.

Facebook Using the Facebook search functionality, we searched for Type 1 diabetes and in doing so identified the largest Type 1 Diabetes Facebook community available. Known as the Type 1 Diabetes Community this community was established in 2011 and is intended to be an open forum for people who have Type 1 diabetes to talk about anything they want. As of 4 October 2014, it had 36,671 likes and on this same date all wall posts and replies that were posted between 1 and 10 September 2014 were identified by viewing the storyline of historical posts. These posts along with the author and the date of posting were manually extracted for further analysis. Given the particular focus of this online community, the sample of 479 posts were all considered relevant to Type 1 diabetes and therefore included in the categorisation stage.

Diabetes.co.uk Diabetes.co.uk is a global Diabetes community with over 125,000 members spanning all forms of the condition. The Type 1 discussion forum on Diabetes.co.uk was identified through the forum homepage and the list of discussion threads was then filtered based on the start date 1 September 2014. All original posts and replies posted between 1 and 10 September 2014, were identified and manually extracted. The total sample of 713 posts was included in the categorisation stage.

Extraction and screening of relevant posts The output of the extraction and screening of Type 1 diabetes posts are summarised in Table 1, where we see the number of posts extracted from each of the three Social Media, the number of original posts and replies included in the sample and their respective number of authors.

III. RQ1: DO DIABETES TYPE 1 PATIENTS USE DIFFERENT SOCIAL MEDIA PLATFORMS FOR DIFFERENT PURPOSES?

Surprisingly, although Twitter had the highest absolute number of total posts at 1433, these results revealed that it is a noisy source of data compared to other Social Media as only 66 posts were relevant to the condition of Type 1 diabetes. Contrary to expectations, the results also revealed that despite being a popular Social Media, Facebook, with a total of 479 posts, was not the most actively used platform for members to discuss the condition. Instead the discussion forum on Diabetes.co.uk was identified as being the most actively used Social Medium included in the study, with a total of 713 posts during the 10-day period.

The sample of Twitter data contained notably fewer responses to posts (0%), than Facebook (96.6%) and Diabetes.co.uk (94.2%). suggesting that there is a greater degree of two-way communication between users of social networks and discussion forums compared to micro-blogging platforms. Twitter has less developed conversational structures, making it harder for patients to read all related comments. Facebook and Diabetes.co.uk posts include associated comments that are easily found. Facebook and Diabetes.co.uk also provided much richer posts, both in terms of length and structured content, i.e. long chains of comments.

When analysing the data further, it was identified that the community moderator who posted questions from anonymous members of community created all original posts within the Facebook community. Thus creating a degree of uncertainty, as the number of authors contributing to original posts is not available This is in contrast to the Diabetes.co.uk discussion forum and Twitter where any registered member of the site could generate an original post and that 37 and 62 members created an original post, respectively.

It was also revealed that a single post within the Facebook community generated a higher response rate relative to others. The post Over/Under time again. 153. Are you over or under? was a request from the community moderator for members to post their current blood glucose levels. This post generated 101 responses, accounting for 21% of the total sample therefore performing a role similar to that of an online survey.

In summary, our results for RQ1 indicate that patients do use different Social Media platforms for different purposes as is highlighted by the volume and conversation structures represented in the sample. Whilst Facebook appeared to be heavily moderated, this was weakly present in Diabetes.co.uk and absent in Twitter. Given these differences in utility it is natural to ask how members of the Type 1 diabetes community use these Social Media to interact with others. Surprisingly, we find that Diabetes.co.uk was the most actively used Social Medium in terms of volume, whilst Facebook achieved the highest percentage response rate. A finding widely known within the Computer Science community but not yet reflected in much of the published Public Health research that is available.

TABLE I. SUMMARY OF POSTS FROM 1 – 10 SEPTEMBER 2014

Social Media	**Sample size**	**Screened sample size**				**Posts included in IPA**
		N (%) original posts	***N authors***	***N (%) replies***	***N reply authors***	
Twitter	1433	66 (4.6%)	62	0 (0%)	0	66
Facebook	479	16 (3.3%)	1	463 (96.6%)	310	479
Diabetes.co.uk	713	41 (5.7%)	37	672 (94.2%)	123	713

IV. RQ2: WHICH SOCIAL MEDIA ARE SUCCESSFUL AT ENCOURAGING SOCIAL INTERACTION AND SUPPORT FOR PATIENTS?

The results of Bales IPA reveal differences in the nature of interactions between users of these three Social Media. These are described in Table II, where we see the percentage of posts relevant to the Bales IPA categories for each of the three Social Media. The Over/Under post accounted for one fifth of the Facebook sample. It was therefore highlighted as a separate line item so as to avoid potential skew of results within the Facebook dataset.

Whilst the three Social Media are predominantly used to disseminate suggestions, opinions and information with other members, the highest percentage of posts; Twitter (36%), Facebook (44%) and Diabetes.co.uk (55%) represent members sharing their opinion. Noticeably fewer posts ask to receive suggestions, opinions and information from other members and the majority of these are original posts.

We observe several interesting differences in the nature of the posts. As noted, Twitter is mainly used to disseminate information (29%) and opinion (36%) and not for interaction. These posts are characterised by dramatization (18%), few friendly posts (18%) and no indications of agreement or disagreement between members. Below shows some examples of these Tweets.

- 33k kids in canada went #backtoschool with diabetes. it’s time to make school a better places for t1d kids.
- sanofi launches mobile game for kids with type 1 diabetes in the uk
- did you know that the character elsa from the movie ”frozen” was in part inspired by a child with type 1 diabetes?
- amazing revolution - bionic pancreas which will automatically inject insulin to type 1 diabetes patients
- an open letter to teresa may advice on her type one diabetes

In contrast, the Type 1 Diabetes Facebook community and Type 1 Diabetes.co.uk discussion forum were considered very interactive with friendly posts accounting for 47% of Facebook posts and 46% of posts on Diabetes.co.uk. The Facebook community generated a greater percentage of agreement (12%) and disagreement (5%) compared to the other Social Media. Whilst the posts within the Diabetes.co.uk discussion forum, indicate a higher percentage of tension (12%) and unfriendly posts (3%), particularly in relation to topics such as diet and the new treatments that are available. Limited moderation of this forum has, therefore, enabled a greater diversity of opinion to be represented.

The three most popular discussion threads on Facebook community are listed in below, including the number of replies to the original post.

- Over/Under time again... 153 Are you over or under? (101)
- This may seem like an odd question but I’m more than a little curious if other T1s experience what I do. I get bit by mosquitos all the time. So much more than anyone else I know. Everyone jokes that mosquitos must love me cause my blood is so sweet. I laugh it all off cause it seems ridiculous. However, I also seem to attract bees. Kind of odd huh? I’m curious how many others experience this...if any? (63)
- Just want peoples advice really I’m 22yrs old been t1 diabetic since I was 6trs old I have one child but planning another but can’t get my hba1c past 8.5 and the docs won’t let me try till it’s 7 any tips on how to get it down? I’m on injections novo rapid and levemir, thanks (54)

The three most popular discussion threads on Diabete.co.uk discussion forum are listed in below, including the number of replies to the original post.

- How highly would you recommend eating low carb? I know this is probably a silly question as it has quite an obvious answer! I love my carbs. I love pasta, rice and potatoes. In the past, I've tried to take the right amount of insulin to cover this but it's so easy to get it wrong and misjudge it - it also means I can end up taking whopping amounts of insulin! Would you recommend I reduce my carbs? It should make my diabetes easier to manage, yes? (I'm also doing Slimming World so although I've read a little about LCHF, I'm not keen to start eating loads of 'fattier' foods!) (91)

- New Flash Glucose Monitoring from Abbott Bloodless Testing Its arrived and heres a video for all you guys who wanted more info (76)
- LCHF success stories from type 1's I created this thread as a place for fellow type 1 diabetics to share their success stories on the LCHF diet. I know there is a similar thread on the low carb forum but I found that most of the responses were from type 2's, so I thought it'd be nice to have a specific place for us to share experiences and hopefully inspire and learn from each other! (67)

In summary, our results indicate that although the three Social Media are all used to disseminate information about the condition, Facebook and Diabetes.co.uk are also used for social interaction and peer support. These findings provoke controversy as to the validity and application of Twitter as a popular Social Media for gaining insight into Type 1 diabetes and in its use as a means of delivering relevant health interventions.

V. RQ3: WHAT ARE THE IMPLICATIONS FOR SOCIAL MEDIA DESIGN?

Although the three Social Media can be considered a valid source of information about the clinical condition of Type 1 diabetes these platforms should not be considered equal or synonymous as has been the case in previous public health studies. Exciting implications are also revealed into the utility of condition specific Social Media as a potentially more effective means of health promotion and patient engagement.

Studies into the social shaping of technology reveal that technology does not develop according to an inner technical logic but is instead a social product influenced by the conditions of its creation and use [7]. With this in mind, we find that the way Social Media are configured and moderated as well as, through the opportunities that they offer for certain types of interaction [8] can shape the behaviours of Social Media users.

For patients living with chronic and life threatening conditions such as diabetes, different utilities are derived from dissemination directed Social Media such as Twitter than from interactive and community building Social Media such as Facebook and Diabetes.co.uk. These differences in use within healthcare are contrary to the generalisations made in existing public health studies, which extrapolate the findings from a single Social Media. This therefore provides interesting and novel applications for a diverse range of research directions that aim to understand how Social Media are used by patients with other clinical conditions and what effect this has on the use of Social Media for research, health interventions and patient engagement.

TABLE II. APPLICATION OF BALES IPA TO DIABETES POSTS

Social Media	**Bales IPA Categories**					
	Shows tension	***Dramatises***	***Agrees***	***Gives Suggestions***	***Gives Opinion***	***Seems Unfriendly***
Twitter	8	18	0	20	36	0
Facebook	4	5	9	29	33	1
Facebook (minus the over/under post)	4	6	12	37	40	1
Diabetes.co.uk	12	7	5	18	55	3
	Gives info	***Asks for info***	***Disagrees***	***Asks for Suggestions***	***Asks for Opinions***	***Seems friendly***
Twitter	29	8	0	5	6	18
Facebook	33	6	4	1	3	37
Facebook (minus the over/under post)	15	8	5	1	3	47
Diabetes.co.uk	32	10	1	1	6	46

VI. CONCLUSION

We presented the first results from this unique study on how different types of Social Media are used by patients living with a chronic condition. In doing so we defy popular assumption and conclude that Type 1 diabetes patients use different Social Media platforms for different purposes, with Twitter primarily used by members for information and opinion sharing, with little support or empathy. Whilst Diabetes.co.uk and Facebook, by virtue of their user base, design and self moderating communities are more successful in their utility for social interaction and peer support by those living with this live long condition. These findings have important implications for Social Media and their application in the context of healthcare.

REFERENCES

[1] M. Hamm, A. Chisholm, J. Shulhan, A. Milne, S. Scott, L. Given, and L. Hartling, “Social media use among patients and caregivers: a scoping review,” BMJ OPEN, vol. 3, no. 5, 2013.

[2] D. Capurro, K. Cole, M. I. Echavarria, N. Joe, J., T., and A. M. Turner, “The use of social networking sites for public health practice and research: A systematic review,” Journal of Medical Internet Research, vol. 16, no. 3, 2014.

[3] S. Moorhead, D. Hazlett, L. Harrison, J. Carroll, A. Irwin, and C. Hoving, “A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication,” Journal of Medical Internet Research, vol. 15, no. 4, 2013.

[4] E. Weitzman, B. Adida, S. Kelemen, and K. Mandl, “Sharing data for public health research by members of an international online diabetes social network,” PlOS One, vol. 6, no. 4, 2011.

[5] L. Ellis, C. Showell, and P. Turner, “Social media and patient self management: not all sites are created equal,” Studies In Health Technology And Informatics, 2013, pp. 291–295.

[6] R. Bales, Interaction process analysis : a method for the study of small groups. Addison-Wesley Press, 1951.

[7] R. Williams and D. Edge, The Social Shaping of Technology. BV, 1996.

[8] P. Smart and N. R. Shadbolt, “Social Machines,” in Encyclopedia of Information Science and Technology. IGI Global, 2014, pp. 6855–6862.

ISBN: 978-1-61208-424-4

A.1.4 British Computing Society Health Informatics Scotland Conference, UK (2017)

Date: 3-4 October 2017

Location: Edinburgh, UK

Type of presentation: Poster and research presentation

A.1.4.1 Abstract submitted to conference

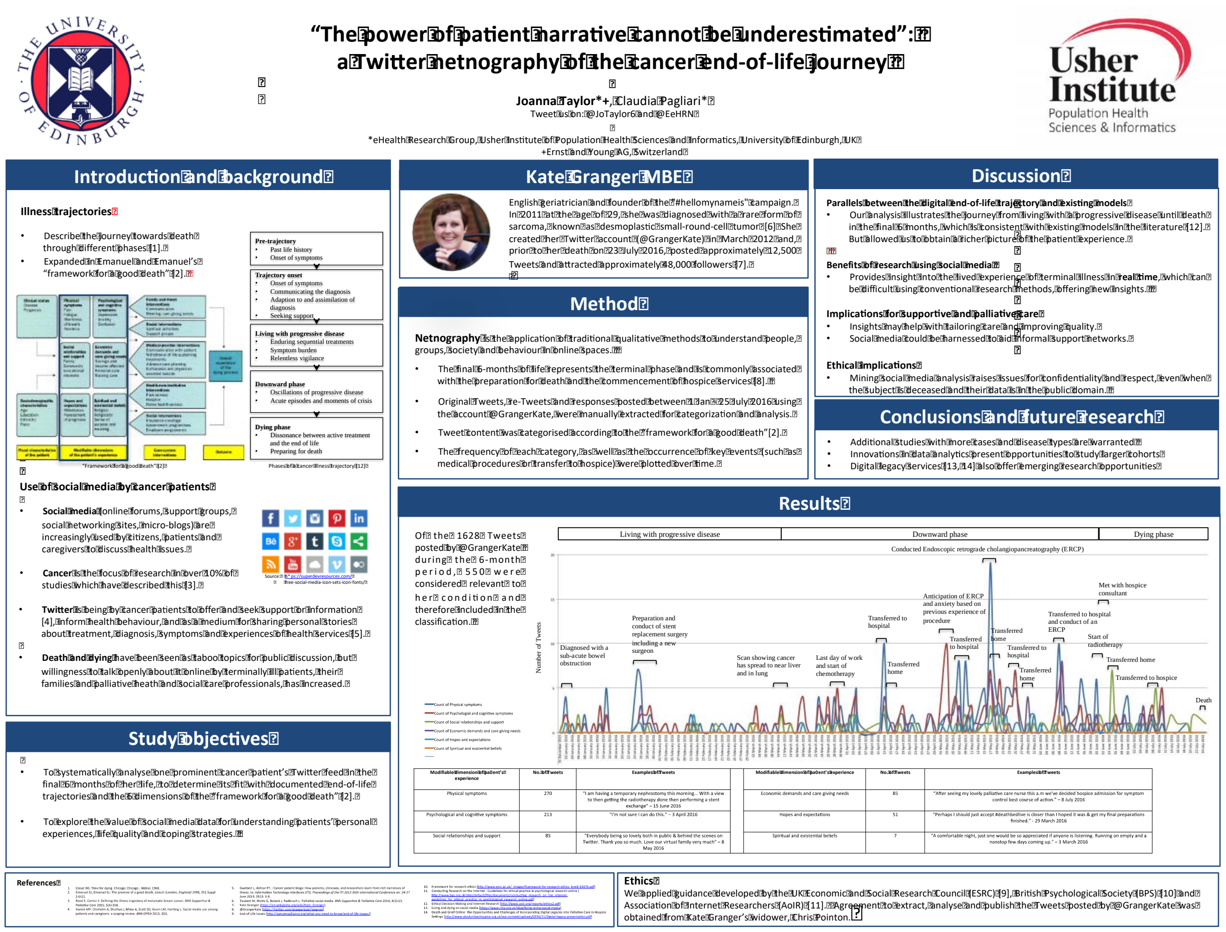
Title: ’*The power of patient narrative cannot be underestimated*’: a Twitter netnography of the cancer end-of-life journey.

Abstract: Interest in the science and practice of palliative care has never been greater, but research to understand the lived experiences of terminally ill patients remains challenging. The growing use of social media offers opportunities to discreetly study the experiences of people facing the journey from diagnosis to death.

Netnography is an application of traditional qualitative methods to understand people, groups, society and behaviour in online spaces. With the consent of her widower, we qualitatively analysed the Twitter feed of prominent cancer patient Dr Kate Granger over the final 6 months of her life.

Drawing on existing frameworks for describing the end-of-life trajectory and a ‘good death’, our analysis revealed clear evidence of the theorised stages of illness and helped to delineate physical, emotional and existential concerns and coping strategies. It also provided unique insights into the benefits of supportive palliative care and online social networks for maintaining psychological resilience and suggests potential opportunities for palliative care professionals and research communities to harness these emerging media and new forms of data. Work in our group is ongoing to analyse a range of comparable cases, using a broader set of analytical tools, as a means of strengthening this novel evidence base.

A.1.4.2 Poster



"The power of patient narrative cannot be underestimated": a Twitter netnography of the cancer end-of-life journey

Joanna Taylor*+, Claudia Pagliari*
Tweet us on: @JoTaylor6 and @EeHRN

*eHealth Research Group, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, UK
+Ernst and Young AG, Switzerland

Introduction and background

Illness trajectories

- Describe the journey towards death through different phases [1].
- Expanded in Emanuel and Emanuel's "framework for a good death" [2].

"Framework for a good death" [2]

Pre-trajectory
- Past life history
- Onset of symptoms

Trajectory onset
- Onset of symptoms
- Communicating the diagnosis
- Adaption to and assimilation of diagnosis
- Seeking support

Living with progressive disease
- Enduring sequential treatments
- Symptom burden
- Relentless vigilance

Downward phase
- Oscillations of progressive disease
- Acute episodes and moments of crisis

Dying phase
- Dissonance between active treatment and the end of life
- Preparing for death

Phases of a cancer illness trajectory [12]

Use of social media by cancer patients

- **Social media** (online forums, support groups, social networking sites, micro-blogs) are increasingly used by citizens, patients and caregivers to discuss health issues.
- **Cancer** is the focus of research in over 10% of studies which have described this [3].
- **Twitter** is being by cancer patients to offer and seek support or information [4], inform health behaviour, and as a medium for sharing personal stories about treatment, diagnosis, symptoms and experiences of health services [5].
- **Death and dying** have been seen as taboo topics for public discussion, but willingness to talk openly about it online by terminally ill patients, their families and palliative heath and social care professionals, has increased.

Source: https://superdevresources.com/free-social-media-icon-sets-icon-fonts/

Study objectives

- To systematically analyse one prominent cancer patient's Twitter feed in the final 6 months of her life, to determine its fit with documented end-of-life trajectories and the 6 dimensions of the "framework for a good death" [2].
- To explore the value of social media data for understanding patients' personal experiences, life quality and coping strategies.

Kate Granger MBE

English geriatrician and founder of the "#hellomynameis" campaign. In 2011 at the age of 29, she was diagnosed with a rare form of sarcoma, known as desmoplastic small-round-cell tumor [6] She created her Twitter account (@GrangerKate) in March 2012 and, prior to her death on 23 July 2016, posted approximately 12,500 Tweets and attracted approximately 48,000 followers [7].

Method

Netnography is the application of traditional qualitative methods to understand people, groups, society and behaviour in online spaces.

- The final 6-months of life represents the terminal phase and is commonly associated with the preparation for death and the commencement of hospice services [8].
- Original Tweets, re-Tweets and responses posted between 1 Jan – 25 July 2016 using the account @GrangerKate, were manually extracted for categorization and analysis.
- Tweet content was categorised according to the "framework for a good death" [2].
- The frequency of each category, as well as the occurrence of key events (such as medical procedures or transfer to hospice) were plotted over time.

Discussion

Parallels between the digital end-of-life trajectory and existing models
- Our analysis illustrates the journey from living with a progressive disease until death in the final 6 months, which is consistent with existing models in the literature [12]. But allowed us to obtain a richer picture of the patient experience.

Benefits of research using social media
- Provides insight into the lived experience of terminal illness in **real time**, which can be difficult using conventional research methods, offering new insights.

Implications for supportive and palliative care
- Insights may help with tailoring care and improving quality.
- Social media could be harnessed to aid informal support networks.

Ethical implications
- Mining social media analysis raises issues for confidentiality and respect, even when the subject is deceased and their data is in the public domain.

Conclusions and future research

- Additional studies with more cases and disease types are warranted
- Innovations in data analytics present opportunities to study larger cohorts
- Digital legacy services [13, 14] also offer emerging research opportunities

Results

Of the 1628 Tweets posted by @GrangerKate during the 6-month period, 550 were considered relevant to her condition and therefore included in the classification.

Modifiable dimension of patient's experience	No. of Tweets	Examples of Tweets
Physical symptoms	270	"I am having a temporary nephrostomy this morning... With a view to then getting the radiotherapy done then performing a stent exchange" – 15 June 2016
Psychological and cognitive symptoms	213	"I'm not sure I can do this." – 3 April 2016
Social relationships and support	85	"Everybody being so lovely both in public & behind the scenes on Twitter. Thank you so much. Love our virtual family very much" – 8 May 2016
Economic demands and care giving needs	85	"After seeing my lovely palliative care nurse this a.m we've decided hospice admission for symptom control best course of action." – 8 July 2016
Hopes and expectations	51	"Perhaps I should just accept #deathbedlive is closer than I hoped it was & get my final preparations finished." - 29 March 2016
Spiritual and existential beliefs	7	"A comfortable night, just one would be so appreciated if anyone is listening. Running on empty and a nonstop few days coming up." – 3 March 2016

References

1. Glaser BG: Time for dying. Chicago: Chicago : Aldine; 1968.
2. Emanuel EJ, Emanuel LL: The promise of a good death. *Lancet (London, England)* 1998, 351 Suppl 2:SII21.
3. Reed E, Corner J: Defining the illness trajectory of metastatic breast cancer. *BMJ Supportive & Palliative Care* 2015, 5(4):358.
4. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, Hartling L: Social media use among patients and caregivers: a scoping review. *BMJ OPEN* 2013, 3(5).
5. Gualtieri L, Akhtar FY, : Cancer patient blogs: How patients, clinicians, and researchers learn from rich narratives of illness. In: *Information Technology Interfaces (ITI), Proceedings of the ITI 2013 35th International Conference on: 24-27 June 2013*; 2013: 3-8.
6. Taubert M, Watts G, Boland J, Radbruch L: Palliative social media. *BMJ Supportive & Palliative Care* 2014, 4(1):13.
7. Kate Granger [https://en.wikipedia.org/wiki/Kate_Granger]
8. @GrangerKate [https://twitter.com/grangerkate?lang=en]
9. End-of-Life Issues [http://sarcomaalliance.org/what-you-need-to-know/end-of-life-issues/]
10. Framework for research ethics [http://www.esrc.ac.uk/_images/framework-for-research-ethics_tcm8-33470.pdf]
11. Conducting Research on the Internet - Guidelines for ethical practice in psychological research online [http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf]
12. Ethical Decision-Making and Internet Research [http://www.aoir.org/reports/ethics2.pdf]
13. Living and dying on social media [https://www.cilip.org.uk/blog/living-dying-social-media]
14. Death and Grief Online: The Opportunities and Challenges of Incorporating Digital Legacies into Palliative Care in Hospice Settings [http://www.stcolumbashospice.org.uk/wp-content/uploads/2016/11/Digital-legacy-presentation.pdf]

Ethics

We applied guidance developed by the UK Economic and Social Research Council (ESRC) [9], British Psychological Society (BPS) [10] and Association of Internet Researchers (AoIR) [11]. Agreement to extract, analyse and publish the Tweets posted by @GrangerKate was obtained from Kate Granger's widower, Chris Pointon.

A.1.5 World Cancer Congress, USA (2018)

Date: 18-20 May 2018

Location: Philadelphia, USA

Type of presentation: Research presentation

A.1.5.1 Abstract submitted to conference

Title: #Deathbedlive: the end-of-life trajectory, reflected in a cancer patient's tweets

Abstract: Interest in the science and practice of palliative care has never been greater but research to understand the lived experiences of terminally ill patients remains challenging. The growing use of social media offers opportunities to discreetly study the experiences of people facing the journey from diagnosis to death.

Netnography is an application of traditional qualitative methods to understand people, groups, society and behaviour in online spaces. With the consent of her widower, we qualitatively analysed the Twitter feed of prominent cancer patient Dr Kate Granger over the final six months of her life. Drawing on existing frameworks for describing the end-of-life trajectory and a 'good death', our analysis revealed clear evidence of the theorised stages of illness and helped to delineate physical, emotional and existential concerns and coping strategies. It also provided unique insights into the benefits of supportive palliative care and online social networks for maintaining psychological resilience and suggests potential opportunities for palliative care professionals and research communities to harness these emerging media and new forms of data. Work in our group is ongoing to analyse a range of comparable cases as a means of strengthening this novel evidence base.

A.2 Peer-reviewed journal manuscripts

The PhD research project described in this thesis was conducted over a six-year period. The research studies described in Chapters 2.2, 4, 7 and 8 addressing RQs 1, 3, 6 and 7, respectively, were published in peer-reviewed journals.

A.2.1 BMJ Open (published December 2018)

Open access Protocol

BMJ Open

Comprehensive scoping review of health research using social media data

Joanna Taylor, Claudia Pagliari

To cite: Taylor J, Pagliari C. Comprehensive scoping review of health research using social media data. *BMJ Open* 2018;**8**:e022931. doi:10.1136/bmjopen-2018-022931

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-022931).

Received 16 March 2018
Revised 29 July 2018
Accepted 11 October 2018

Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK

Correspondence to
Dr Claudia Pagliari;
claudia.pagliari@ed.ac.uk

ABSTRACT

Introduction The rising popularity of social media, since their inception around 20 years ago, has been echoed in the growth of health-related research using data derived from them. This has created a demand for literature reviews to synthesise this emerging evidence base and inform future activities. Existing reviews tend to be narrow in scope, with limited consideration of the different types of data, analytical methods and ethical issues involved. There has also been a tendency for research to be siloed within different academic communities (eg, computer science, public health), hindering knowledge translation. To address these limitations, we will undertake a comprehensive scoping review, to systematically capture the broad corpus of published, health-related research based on social media data. Here, we present the review protocol and the pilot analyses used to inform it.

Methods A version of Arksey and O'Malley's five-stage scoping review framework will be followed: (1) identifying the research question; (2) identifying the relevant literature; (3) selecting the studies; (4) charting the data and (5) collating, summarising and reporting the results. To inform the search strategy, we developed an inclusive list of keyword combinations related to social media, health and relevant methodologies. The frequency and variability of terms were charted over time and cross referenced with significant events, such as the advent of Twitter. Five leading health, informatics, business and cross-disciplinary databases will be searched: PubMed, Scopus, Association of Computer Machinery, Institute of Electrical and Electronics Engineers and Applied Social Sciences Index and Abstracts, alongside the Google search engine. There will be no restriction by date.

Ethics and dissemination The review focuses on published research in the public domain therefore no ethics approval is required. The completed review will be submitted for publication to a peer-reviewed, interdisciplinary open access journal, and conferences on public health and digital research.

Strengths and limitations of this study

- ► The proposed scoping review addresses the need for a comprehensive, cross-disciplinary synthesis of health-related research using social media data.
- ► A five-stage scoping review protocol was developed, which was informed by an exploratory analysis of existing reviews, terminologies and published taxonomies.
- ► This revealed the changing usage of relevant terms over time, the types of keyword combinations most likely to yield eligible studies and the time-lag between the launch of new social media platforms and published research using data derived from them.
- ► The findings of the scoping review will be communicated using both static and multimedia visualisation tools.

INTRODUCTION

Social media are online, often mobile, platforms that support the creation and exchange of user-generated content.[1] They include generic platforms for networking, information sharing and collaboration (eg, Facebook, Twitter, YouTube, Linkedin) and online forums aimed at specific communities (eg, Patientslikeme, Mumsnet). The first modern day social media platform, sixdegrees, was launched in 1997 and allowed users to connect with friends and family through sending messages and posts on bulletin boards.[2] Since then the use of social media has become increasingly common[3] with an estimated 2.5 billion of the global population estimated to be using them as of 2017.[4]

As social media have become an integral part of people's lives, research focusing on social media has also evolved and is taking place in a number of domains. A bibliometric content analysis of key words that appeared in the abstracts of 14 500 journal articles and conference papers identified five main areas in which a high volume of research already exists. This includes research related to health, educational uses of social media, computing and computer science methods, business, organisational and marketing topics and political and social engagement.[5]

Uses of social media in the context of research can be described according to the broad categories identified by Bjerglund-Andersen and Söderqvist (dissemination, discussion/networking, public engagement, teaching, research/data collection).[6] Taylor and Pagliari[7] [7] recently separated the last of these into social media as a source of data for research and as

Table 1 Existing systematic, quasi-systematic or scoping reviews indexed in PubMed

Topic	Examples of systematic, quasi-systematic or scoping reviews
Disease surveillance	Social media and internet-based data in global systems for public health surveillance: a systematic review.[16] Scoping review on search queries and social media for disease surveillance: a chronology of innovation.[17] Ebola and the social media.[18] Digital disease detection: a systematic review of event-based internet biosurveillance systems.[19] Utility and potential of rapid epidemic intelligence from internet-based sources.[20] Using online social networks to track a pandemic: a systematic review.[21] A systematic review of event-based public health surveillance systems.[22] Social media: a systematic review to understand the evidence and application in infodemiology.[23]
Adverse event monitoring	Systematic review on the prevalence, frequency and comparative value of adverse events data in social media.[26]
Quality of healthcare services	Social media and rating sites as tools to understanding quality of care: a scoping review.[24] Eight questions about physician-rating websites: a systematic review.[25]
Illicit drug use	Systematic review of surveillance by social media platforms for illicit drug use.[27]
eGovernment	Use of social media for e-Government in the public health sector: a systematic review of published studies.[28]
Chronic disease	Social media use in chronic disease: a systematic review and novel taxonomy[29]
Ethics	Attitudes toward the ethics of research using social media: a systematic review[30]

a tool for the conduct of research. These ‘secondary uses’ of social media data include analyses of trends, associations and sentiments in users’ postings, as well as interactions and networks.[8 9]

This new source of ‘big data’ has triggered scientific developments in a number of areas, including health. The term ‘infoveillance’ was coined by Eysenbach in 2009 to describe the automated harvesting and analysis of internet searches and social media postings as an alternative approach for health and disease surveillance.[10] Other terms have since been introduced across academia and industry to describe the use of such data for gleaning insights about behavioural trends, determining the impact of interventions or predicting future events. These include ‘social media listening’,[11] ‘social media mining’,[12] ‘social analytics’,[13] ‘social machines’[14] and ‘netnography’,[15] to name but a few.

Based on an iterative search using PubMed, we identified an existing corpus of 15 systematic, quasi-systematic and scoping reviews on the secondary use of social media data for health research, examples of which are summarised in table 1 (see online supplementary appendix 1 for the

Table 2 Research questions

Aspect	List of questions
General	What is the total number of studies published by year? What terms are being used to describe the nature of this research? Which academic communities are most active in health research using data from social media? Where are study authors located, according to their affiliation? What is the geographical scope of the social media data analysed in these studies? For what purposes are social media data being used in this research?
Topic	Which health topics are being studied?
Social Media type	Which social media platforms or sites are being used as sources of data?
Extract and analysis	What units of analysis are being applied? How are data from social media being extracted and analysed and which proprietary tools are being used?
Ethics	How are ethical considerations applied in the published research?

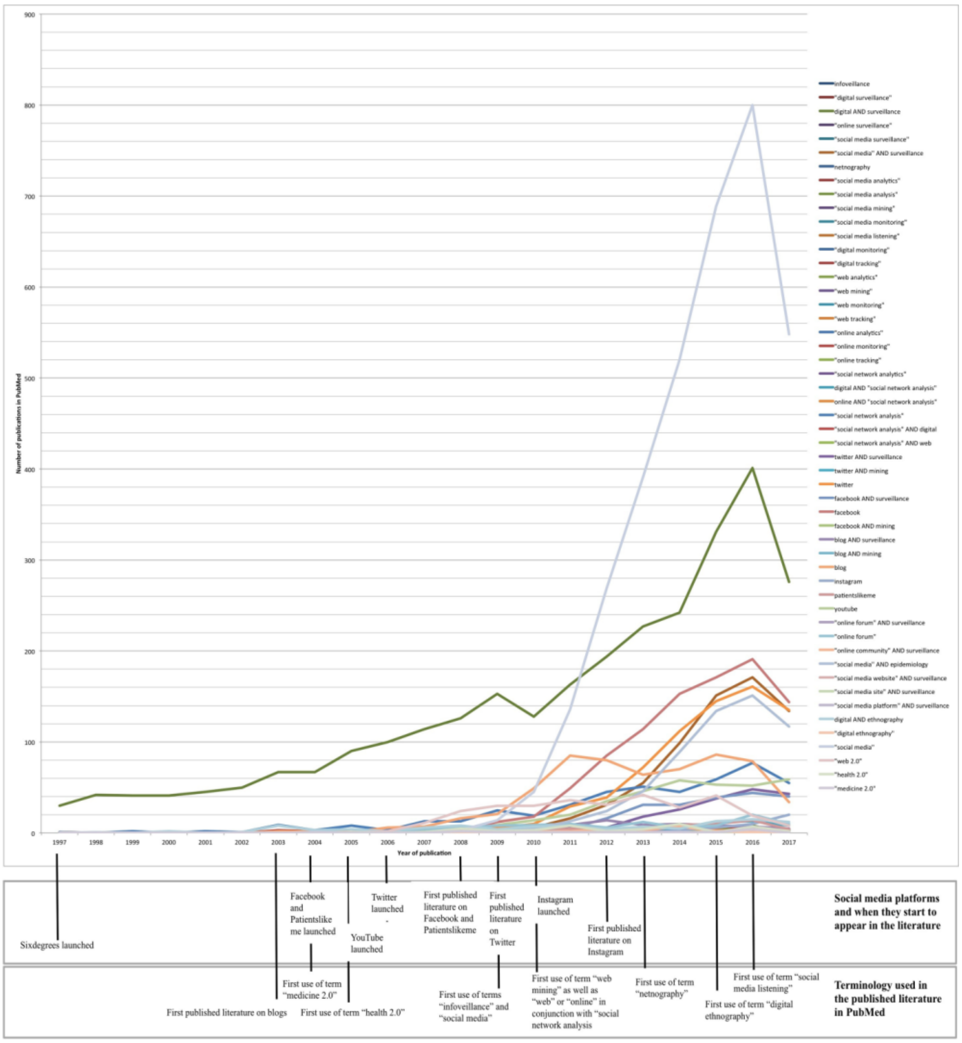


Figure 1 Results of the search queries by year.

full list, summarised according to their aims, number of databases and findings). Among these are reviews on approaches to health surveillance[16–23] quality of healthcare services,[24 25] adverse event monitoring,[26] illicit drug use,[27] eGovernment,[28] chronic conditions[29] and ethics.[30]

A 2013 scoping review by Bernardo *et al* on the use of internet search queries and social media for disease surveillance identified the first study as has having been published in 2006 and described how techniques for exploiting this data are evolving to increase the accuracy of signal detection.[17] Although insightful in understanding the chronology of this type of innovation, the review searched only one database and focused on the surveillance of foodborne illness. A 2014 systematic review by Velasco *et al* on the impact and use of data from social media for public health surveillance went further, concluding that existing surveillance systems are limited, and there is a need for automated technologies to monitor health-related information on the internet,

Box 1 Article inclusion and exclusion criteria for the scoping review

Inclusion

- Types of publication: peer-reviewed research articles. Full conference papers.
- Language: English.
- Publication timeframe: 1997–2017.
- Types of research: empirical studies using health-related data from social media, extracted or studied in situ, using both manual and automated methods.

Exclusion

- Types of publication: dissertations/theses; reports or abstracts only; letters to the editor; marketing or advertising material; reviews or editorials.
- Language: not English.
- Types of research: studies based on data from online sources other than social media (eg, internet search histories, online news reports). Commercial research aimed at obtaining market intelligence or informing product promotion. Studies examining social media platforms, rather than using them as a source of data. Studies describing social media as a communication or broadcasting channel (eg, for public health promotion).

although it did not specify which types of systems were analysed.[16] O'Shea's systematic review of event-based internet biosurveillance systems describes the wide variety of technologies and data sources for gathering, processing and disseminating data to detect infectious disease outbreaks.[19] These reviews, alongside the others listed, focus predominantly on infectious disease. A systematic review on social media for chronic disease exists[29] which focuses on understanding the clinical outcomes associated with using such technologies for patient support, education and disease management across different conditions. However, this is limited by the inclusion of only one database and did not examine the type of methods or tools used to extract and analyse social media data, the academic discipline and setting of the research or the ethical issues considered. With respect to the latter, a recent systematic review by Golder *et al*, analysed studies reporting attitudes towards the ethics of research using social media data. This revealed wide variation in attitudes, from the very positive to the very concerned, depending on the purpose and quality of the research, researcher affiliation, the potential for harm and the methods used.[30] Although it used an impressive 16 databases, this review did not examine regional, disciplinary, condition or health topic specific variations, and the authors note that the demographic characteristics of respondents were unclear in most studies. As noted in a recent review of UK research ethics guidelines and published health research, growing public awareness of the misuse of social media for marketing and algorithmic prediction are forcing policymakers to look more closely at this issue.[7]

In light of these gaps, we set out to undertake a comprehensive scoping review aimed at capturing and profiling a broad corpus of published multidisciplinary research in which data obtained from social media have been used to monitor, understand or evaluate aspects of health and disease. In this paper, we describe the formal protocol for the comprehensive scoping review, alongside the preliminary analyses undertaken to inform each stage.

METHODS AND ANALYSIS

Scoping reviews are a type of quasi-systematic review that are increasingly used for understanding research on emerging innovations, which may be poorly indexed, distributed across published and grey literature or located in different academic disciplines.[31–33] They typically progress in five key stages (1) identifying the research question; (2) identifying the relevant studies; (3) selecting the studies; (4) charting the data and (5) collating, summarising and reporting the results. Developing an a priori review protocol can be useful for managing this complexity, while formative research can aid the design of such protocols by identifying relevant terminologies, topics and evidence sources. Scoping reviews are mainly aimed at mapping the evidence landscape rather than establishing the effectiveness of particular interventions, and typically do not involve critical appraisal of study methodology or detailed extraction of outcomes data.[34]

Stage 1: identifying the research question

In addition to the overarching review objective articulated in the introduction, several specific questions will be used to guide our analysis of existing research evidence, as listed in table 2.

Stage 2: identifying relevant literature

Comprehensive scoping reviews aim to capture literature from a range of electronic databases, reference lists and grey literature. As such, our approach will include:

- A systematic search of peer-reviewed studies using five health, informatics, business and cross-disciplinary electronic databases: PubMed, Scopus, Association of Computer Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE) and Applied Social Sciences Index and Abstracts (ASSIA).
- 'Snowballing' from article reference lists[35] will be used to identify additional studies that may have not been indexed in the online research databases.
- Searching grey literature from the internet using the most widely used search engine, Google.[36] The first 20 Google results yielded by each search string will be reviewed, as this search engine displays the results by relevance.

The search strategy for each of the databases was defined in consultation with a senior librarian. Our five electronic databases and one internet search engine takes account of time and funding constraints, although the sources targeted are likely to capture most of the relevant literature.[37]

Formative searches

Given the breadth and changing popularity of terms related to social media and social media mining, we undertook a formative analysis to understand those most likely to yield articles relevant to our review objectives. We first created an extensive list based on the search strings specified in a recent systematic review of social media in the context of e-Government in public health,[28]

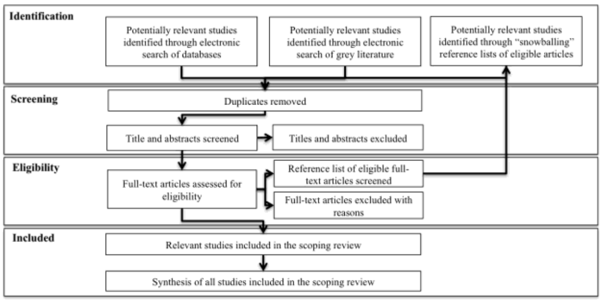


Figure 2 Steps for identification, screening, eligibility and inclusion of studies in the scoping review.

supplemented with terms captured from iterative online searches. The list included generic terms related to social media (eg, digital, blog, social network, Web 2.0), named social media (eg, Twitter, Facebook), types of social media mining (eg, surveillance, scraping, listening, infoveillance) and analytics (eg, netnography, social network analysis) (see online supplementary appendix 2 for the full list). We then ran separate search queries for each term in PubMed, constructed as [term] AND (health OR illness OR disease) AND human. Searches yielding one or more hits were charted over time, to examine their changing frequency in the research literature. Out of the 72 terms tested in PubMed, 52 produced hits. The timeline was cross-referenced with key historical events, such as the introduction of new platforms or new methodologies. The results are shown in figure 1.

As can be seen from figure 1, 91% of relevant studies were published in the past 9 years (11 154 results between 2009–2017 and 1048 between 1997–2008). Not surprisingly, the search term 'social media' produced the highest number of results overall compared with other search queries. Regarding specific platforms, searches with sixdegrees and Linkedin produced no results, compared with other social media such as Facebook (1152 hits, between 2008 and 2017), Twitter (939 hits, between 2009 and 2017), YouTube (351 hits, between 2008 and 2017), Patientslikeme (60 hits, between 2008 and 2017) and Instagram (40 hits, between 2012 and 2017). We also observed a time lag of 4 years between when Facebook, Patientslikeme and Twitter were launched in 2004, 2004 and 2006 respectively and the first research article was published, while the equivalent latency for YouTube and Instagram was 2 years. The individual search terms 'surveillance' (665), 'epidemiology' (581) and 'ethnography' (110) produced the highest yields, compared with terms specific to digital research methods. With respect to the latter, temporal changes in the use of terms such as 'infoveillance', 'netnography', 'social media listening', 'social media analytics' and 'social media mining' indicate the evolution of innovations and research perspectives, however including such specialist terms was not critical for identifying relevant studies.

As previously noted, a systematic search of five health, informatics, business and cross-disciplinary electronic databases—PubMed, Scopus, ACM, IEEE and ASSIA—will be undertaken to identify relevant peer-reviewed studies. Below is the search query that will be used to interrogate these sources. This takes account of insights gathered during our formative analyses:

- Named social media,[38] including health-specific social media sites.[39]
- OR 'web 2.0' OR 'social media' OR 'blog' OR 'microblog' OR 'wiki' OR 'virtual world' OR 'discussion forum' OR 'online community'
- AND 'surveillance' OR 'epidemiology' OR 'infoveillance' OR 'ethnography' OR 'netnography' OR 'mining'
- AND ('health' OR 'disease' OR 'illness')
- NOT 'animal'

For Google searches the advanced search option will be used with English selected as the preferred language. All regions will be included, and the following search strings will be used:

- social media, surveillance, health, NOT animal'
- 'social media, surveillance, disease, NOT animal'
- 'social media, data mining, health, NOT animal'
- 'social media, data mining, disease, NOT animal'
- 'social media, epidemiology, health, NOT animal'
- 'social media, epidemiology, disease, NOT animal'

Stage 3: study selection

One researcher will conduct the systematic search of electronic databases and grey literature. Studies will be selected after the abstracts and titles, identified via the electronic searches and 'snowballing' technique, have been independently screened for relevance by two researchers based on the specified inclusion criteria.

Inclusion and exclusion criteria

Peer-reviewed journal articles and conference papers will be included where these describe empirical research using data from social media such as Twitter or Facebook, whether extracted or studied situ, using either manual or automated methods. Studies not in English, dissertations/

Table 3 Existing classification frameworks that correspond with the research questions

Research question	Existing classification framework to be applied	How the classification framework will be applied
What is the total number of studies published by year?	Not specified	The year that the eligible study was published will be captured.
What terms are being used to describe the nature of the research?	Not specified	The terminology used by study authors to describe the aims or methodologies used in their study will be captured for example, infoveillance, netnography, social listening.
Which academic communities are most active in conducting health research studies that use data from social media?	Scimago Journal Subject Areas[40] including 27 subject areas including medicine, computer science, health professions, business management and accounting as well as social sciences.	The journal in which eligible study is published, along with the affiliation of all authors as well as any sources of funding for the study (if shown) will be classified according to the disciplinary taxonomy used by the Scimago Journal ranking portal. These categories are not mutually exclusive.
Where is the affiliation of the first author located?	World Bank Regional and Lending Groups for Countries includes seven regions and four lending groups.[41]	The geographical location of the first author conducting the eligible study will be classified according to the regional and lending groups provided by the World Bank for 2017. These categories are not mutually exclusive.
What is the geographical scope of the sample of social media data analysed in the study?	World Bank Regional and Lending Groups for Countries includes seven regions and four lending groups.[41]	The geographical location of the population sample analysed within the study will be classified according to the regional and lending groups provided by the World Bank for 2017. These categories are not mutually exclusive.
What is the purpose for which social media data are being used in the research?	Not specified	The purpose of the eligible study will be captured.
Which health topics are being studied?	The 19 health-related topics which inform the WHO's Sustainable Development Goals (SDG).[42]	The type of condition and health topic being studied in the eligible study will be classified according to the list of health-related topics, which inform the WHO SDG. These categories are not mutually exclusive.
Which social media are used as a source of data?	Kaplan and Haenlein's eight types of social media.[1]	The type of social media from which the data for the eligible study was sourced will be classified according to Kaplan and Haenlein's eight types of social media. The name of the social media will also be captured.
How large are the studies and what is the unit of analysis applied?	Not specified	The sample size and unit of analysis of the eligible study will be captured.
How is the data from social media extracted and analysed and which proprietary tools were used?	Several types of analysis may be performed on social media data, ranging from simple descriptive statistics to qualitative research to automated real-time analytics at scale as described by Batrinca and Treleaven.[43] These types of analysis have been described in terms of computational techniques, such as natural language processing and purposive approaches such as news analytics, opinion mining, data scraping, sentiment analysis and text analytics to name but a few.	Batrinca and Treleaven taxonomy of social media analysis techniques will be used to guide our classification of the broad approaches and specific techniques demonstrated in the eligible studies. This taxonomy may be refined as a result of new insights emerging during data extraction. We will also capture the named type of analysis performed in the eligible study as well as any reference to proprietary tools used.
How are ethical considerations applied in the published research?	Conway's taxonomy of ethics concepts for the use of Twitter in public health surveillance and research[44] and can be applied across all manner of social media.	The application of ethical concepts in the eligible studies will be classified according to the 10 high level categories identified by Conway. These categories are not mutually exclusive.

theses, reports or abstracts, letters to the editor and feature articles and articles intended as marketing or advertising material will be excluded. A publication timeframe of 1997–2017 will be applied. See box 1 for the inclusion and exclusion criteria for the scoping review.

Intermittent cross-checking by the second author will help to ensure the appropriate application of inclusion and exclusion criteria. In a further stage, the full-text of included articles will be independently assessed for eligibility by both authors and discrepancies resolved through discussion. Reasons for exclusion will be documented. These steps are described in figure 2.

EndNote reference management software will be used to manage the records retrieved from searches. One reviewer will independently screen the generated citations with the help of EPPI-Reviewer 4 systematic review software and undertake data extraction.

Stage 4: charting the data

The purpose of charting data in scoping reviews is to produce a descriptive summary of the results. For this stage, we have identified existing classification frameworks that correspond with the research questions listed previously. These will form the basis of our data charting form. The classification frameworks consider the purpose for which social media data are being used in the research, the method of data extraction (including any automated data mining tools used), the analytical–interpretive approach used (including stated theoretical perspectives), the locus of the research by academic institution and geographical scope of the data, the academic discipline associated with the research and whether/how ethical issues or guidelines are considered. Each of these existing classification frameworks and taxonomies are considered in further detail in table 3 in relation to the specific research question. These classification frameworks are not intended to be totally prescriptive, and additional emerging themes will be captured throughout the forthcoming analysis. These themes will later be used to identify gaps and inconsistencies represented in the existing frameworks, for future consideration and refinement.

Stage 5: collating, summarising and reporting the results

The extracted data will be tabulated, with rows relating to articles, columns to classification variables and cells containing the relevant information. Both frequency analysis and trend analysis will be used to chart the classified results.

Frequency analysis

Using frequency analysis, the counts and percentages of eligible studies will be calculated. Studies will be grouped based on the classification frameworks and taxonomies applied. The result of this analysis will be a map of eligible studies represented in bubble plot, graph or tabular form.

Trend analysis

Trend analysis will be used to present the changing frequency of research over the past 20 years, based on the aforementioned classification criteria. A map showing the characteristics of included studies will be presented using both static and multimedia visualisation tools, such as an animated bubble plot or graph.

Public and patient involvement

No patients or members of the public were involved in the protocol design or exploratory analyses, nor do we plan to include them in the conduct of the scoping review. Our results are nevertheless likely to be of interest to citizens who use social media, and our decision to examine how researchers report ethical considerations reflects this concern. Dissemination will include accessible summaries and graphics, which we intend to make available to the public via social media.

ETHICS AND DISSEMINATION

The completed scoping review will be submitted for publication to a peer-reviewed, interdisciplinary open access journal, in addition to conferences on public health and digital research. Findings will be presented using both static and multimedia visualisation tools.

Acknowledgements Thanks to Marshall Dozier for advice on the search strategy and Santosh Vijaykumar for comments on a previous draft of the manuscript.

Contributors JT: designed the protocol and undertook the exploratory searches, with input from CP. JT: drafted the manuscript. CP: edited the manuscript.

Funding JT is a self-funded PhD student at the University of Edinburgh, supervised by CP. She also works as a management consultant for Ernst and Young Switzerland although her employers are not involved in this research. CP is a collaborator on three RCUK-funded research programmes associated with this study: The Administrative Data Research Centre for Scotland, ESRC (Grant number ES/ L007487/1); The Science and Practice of Social Machines EPSRC (EP/ J017728/1) and the Farr Institute for Health Informatics Research (Scotland) MRC (Grant number MR/K007017/1).

Disclaimer The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests None declared.

Patient consent Not required.

Ethics approval Although some scoping reviews include a consultation phase, this one focuses on published research and online sources already in the public domain. No ethics approval is therefore required.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. *Bus Horiz* 2010;53:59–68.
2. Small Business Trends. complete history of social media: then and now. https://smallbiztrends.com/2013/05/the-complete-history-of-social-media-infographic.html
3. Digital in 2017. Global overview. https://wearesocial.com/special-reports/digital-in-2017-global-overview
4. The Statistics Portal. Social media - statistics and facts. https://www.statista.com/topics/1164/social-networks/
5. Social Media and Society International CoSMLSMS Logos JPG v2nference. Current state of social media research: from practice to

theory #SMSociety15. http://socialmedialab.ca/2015/current-state-of-social-media-research-from-practice-to-theory-part-1/

6. Faculty o f Science. Social media and public health research. 2012 https://bjerglund.files.wordpress.com/2012/11/final-social-media-and-public-health-research1.pdf
7. Taylor J, Pagliari C. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? *Research Ethics* 2017:1747016117738559.
8. Anstead N, O'Loughlin B. Social media analysis and public opinion: the 2010 uk general election. *Journal of Computer-Mediated Communication* 2015;20:204–20.
9. The Leading Association of Public Opinion and Survey Research Professionals. Social media in public opinion research: report of the AAPOR task force on emerging technologies in public opinion research. https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf
10. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11:e11.
11. Smith T. The social media revolution. *Int J of MarK Res* 2009;51:559.
12. Zafarani R, Ali Abbasi M, Liu H. *Social media mining*: An Introduction: Cambridge University Press, 2014.
13. IT Glossary. Social Analytics. http://www.gartner.com/it-glossary/social-analytics
14. Smart PR. Global Khosrow-Pour MShadbolt NR: Social Machines. , *Encyclopedia of Information Science and Technology. edn*: IGI Global, 2014.
15 Kozinets RV. Netnography. *The International Encyclopedia of Digital Communication and Society*: John Wiley and Sons, Inc, 2015.
16. Velasco E, Agheneza T, Denecke K, *et al*. Social media and internet-based data in global systems for public health surveillance: a systematic review. *Milbank Q* 2014;92:7–33.
17. Bernardo TM, Rajic A, Young I, *et al*. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. *J Med Internet Res* 2013;15:e147.
18. Fung IC, Duke CH, Finch KC, *et al*. Ebola virus disease and social media: a systematic review. *Am J Infect Control* 2016;44:1660–71.
19. O'Shea J. Digital disease detection: A systematic review of event-based internet biosurveillance systems. *Int J Med Inform* 2017;101:15–22.
20. Yan SJ, Chughtai AA, Macintyre CR. Utility and potential of rapid epidemic intelligence from internet-based sources. *Int J Infect Dis* 2017;63:77–87.
21. Al-Garadi MA, Khan MS, Varathan KD, *et al*. Using online social networks to track a pandemic: A systematic review. *J Biomed Inform* 2016;62:1–11.
22. A systematic literature review on event-based public health surveillance systems. http://edoc.sub.uni-hamburg.de/haw/volltexte/2012/1614/pdf/lsab12_27.pdf
23. Guy S, Ratzki-Leewing A, *et al*. Social Media: A Systematic Review to Understand the Evidence and Application in Infodemiology. *Electronic healthcare: 2012// 2012*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012:1–8.
24. Verhoef LM, Van de Belt TH, Engelen LJ, *et al*. Social media and rating sites as tools to understanding quality of care: a scoping review. *J Med Internet Res* 2014;16:e56.
25. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. *J Med Internet Res* 2013;15:e24.
26. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. *Br J Clin Pharmacol* 2015;80:878–88.
27. Kazemi DM, Borsari B, Levine MJ, *et al*. Systematic review of surveillance by social media platforms for illicit drug use. *J Public Health* 2017;39:763–76.
28. Tursunbayeva A, Franco M, Pagliari C. Use of social media for e-Government in the public health sector: A systematic review of published studies. *Gov Inf Q* 2017;34:270–82.
29. Patel R, Chang T, Greysen SR, *et al*. Social Media Use in Chronic Disease: a systematic review and novel taxonomy. *Am J Med* 2015;128:1335–50.
30. Golder S, Ahmed S, Norman G, *et al*. Attitudes toward the ethics of research using social media: a systematic review. *J Med Internet Res* 2017;19:e195.
31. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8:19–32.
32. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010;5:69.
33. JBI Wiki. Chapter 11: Scoping Reviews. https://reviewersmanual.joannabriggs.org/
34. Armstrong R, Hall BJ, Doyle J, *et al*. Cochrane update. 'scoping the scope' of a cochrane review. *J Public Health* 2011;33:147-50.
35. Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. *Cochrane Database Syst Rev* 2011(8):MR000026.
36. Market Share Statistics for Internet Technologies. Search engine market share. https://netmarketshare.com/search-engine-market-share.aspx
37. Lam MT, McDiarmid M. Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014. *J Med Libr Assoc* 2016;104:284–9.
38. The Free Wikipedia. List of social networking websites. http://en.wikipedia.org/wiki/List_of_social_networking_websites
39. Top 25 health and medical social media sites. http://mastersinhospitaladministration.com/2011/top-25-health-and-medical-social-media-sites/
40. SJR. Scimago journal and country rank. http://www.scimagojr.com/journalrank.php
41. The World Bank. World bank country and lending groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups
42. World Health Statistics. 2016 data visualizations dashboard: monitoring health for SDGs. http://apps.who.int/gho/data/node.sdg.tp-1?lang=en
43. Batrinca B, Treleaven PC. Social media analytics: a survey of techniques, tools and platforms. *AI Soc* 2015;30:89–116.
44. Conway M. Ethical issues in using twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. *J Med Internet Res* 2014;16:e290.

A.2.2 Research Ethics (published October 2017)

Altmetric attention score: 72 (as of 2 May 2020)

Number of citations: 24 (as of 2 May 2020)

Article

Mining social media data: How are research sponsors and researchers addressing the ethical challenges?

Research Ethics
1–39

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1747016117738559
journals.sagepub.com/home/rea

Joanna Taylor
Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, UK
Ernst and Young Ltd, Switzerland

Claudia Pagliari
Usher Institute for Population Health Sciences and Informatics, The University of Edinburgh, UK

Abstract

Background: Data representing people's behaviour, attitudes, feelings and relationships are increasingly being harvested from social media platforms and re-used for research purposes. This can be ethically problematic, even where such data exist in the public domain. We set out to explore how the academic community is addressing these challenges by analysing a national corpus of research ethics guidelines and published studies in one interdisciplinary research area.
Methods: Ethics guidelines published by Research Councils UK (RCUK), its seven-member councils and guidelines cited within these were reviewed. Guidelines referring to social media were classified according to published typologies of social media research uses and ethical considerations for social media mining. Using health research as an exemplar, PubMed was searched to identify studies using social media data, which were assessed according to their coverage of ethical considerations and guidelines.
Results: Of the 13 guidelines published or recommended by RCUK, only those from the Economic and Social Research Council, the British Psychological Society, the International Association of Internet Researchers and the National Institute for Health Research explicitly mentioned the use of social media. Regarding data re-use, all four mentioned privacy issues but

---

Corresponding author:
Claudia Pagliari, eHealth Research Group, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Medical School (Door 3), Teviot Place, Edinburgh EH8 9AG, UK.
Email: claudia.pagliari@ed.ac.uk

varied with respect to other ethical considerations. The PubMed search revealed 156 health-related studies involving social media data, only 50 of which mentioned ethical concepts, in most cases simply stating that they had obtained ethical approval or that no consent was required. Of the nine studies originating from UK institutions, only two referred to RCUK ethics guidelines or guidelines cited within these.

Conclusions: Our findings point to a deficit in ethical guidance for research involving data extracted from social media. Given the growth of studies using these new forms of data, there is a pressing need to raise awareness of their ethical challenges and provide actionable recommendations for ethical research practice.

Keywords

Social media, Internet, ethics, guidelines, data science, digital research

Introduction

Social media in research

Technological advances over the past decade have enabled widespread access to the Internet in most countries and the number of social media users has grown to around 2.8 billion people worldwide (Kemp, 2017). Social media are online, often mobile, platforms that support the creation and exchange of user-generated content (Kaplan and Haenlein, 2010), a phenomenon sometimes referred to by the terms Web 2.0 or the Social Web. They include generic platforms for networking, information sharing and content curation, such as Facebook,[1] Twitter,[2] YouTube[3] and LinkedIn[4]; online forums aimed at specific communities, such as PatientsLikeMe,[5] Mumsnet[6] and BaristaExchange[7]; some private collaborative work tools such as Trello[8] and Yammer[9]; and crowdsourcing platforms such as Ushahidi[10] and Zooniverse[11], although opinions vary as to what precisely does or does not qualify.

Several uses of social media in research have been described in the literature. These include the deployment of social media platforms for the *conduct* of research, such as for gathering opinions (Hilyard et al., 2015), recruiting study participants (Pedersen and Kurz, 2016), undertaking participative ‘citizen science’ (Trisha, 2013) or fostering stakeholder involvement (Russell et al., 2016). People’s online activity in social media is also increasingly being used as a *source of data* for research (Wilson et al., 2012). Such ‘secondary uses’ include studies seeking to profile or understand users’ behaviours, demographics, interactions and networks, or to assess their responses or sentiments towards particular topics, products or policies (Anstead and O’Loughlin, 2015; Murphy et al., 2014). One of the most significant trends, from both a scientific and societal perspective, is the application of automated tools for mining and analysing social media as a means of revealing new associations or predicting future behaviours or outcomes. Increasingly this is taking place alongside data mining from institutional or business repositories, to link historical and real-time information (Smith, 2014). While the business sector has been using social media data for some time; such as to monitor brand reputation; their value for academic research is gradually being realised. In the United Kingdom (UK) considerable government funding has been invested in a network of major 'big data' research centres. Although these are mainly concerned with public sector administrative data (including health, housing and tax records, amongst others), recent investments include research centres focused on social media (Cardiff University, 2012).

A number of potential benefits of using social media in research have been described in the literature, including the ability to reach larger numbers of participants than might otherwise be possible (Moorhead et al., 2013), being able to analyse trends and associations within large corpuses of open-access data (Paul and Dredze, 2011), reducing the costs of conducting research in large populations (Munson et al., 2013), greater opportunities for interaction across extended time periods, as may be required in longitudinal or post-market studies (Hokby et al., 2016), providing a channel for social research that is less prone to bias than approaches involving direct contact between researchers and participants (McKee, 2013), involving citizens in the research process (INVOLVE, 2014), being able to curate and enrich biomedical knowledge (Good et al., 2012) and generating new channels for research dissemination (Balm, 2014).

Methodological and ethical challenges

Despite these advantages, the complexity of interactions between individuals, groups and technical systems in these online spaces presents a number of challenges for academics wishing to use social media data in research (Munson et al., 2013). These include the self-selecting nature of social media users, inequalities in access to social media platforms and data, the difficulty of obtaining meaning from heterogeneous data of variable quality and provenance, and a dependence on observing and interpreting what is 'out there' in a way that differs from traditional sampling approaches. Arguably, however, the greatest challenges for researchers in this area are ethical ones (David, 2004; Eysenbach and Till, 2001), such as variable perceptions of and unclear boundaries between 'public' and 'private' spaces, as well as the difficulty of ensuring anonymity and preserving the privacy of data subjects, whose identities may not be disguised or may be easily deduced from their postings and affiliations. Related issues of ownership and intellectual property are also poorly defined and consent to the use of social media data in research is rarely obtained through informed choice, but rather assumed on the basis that users have chosen to place it in the public domain (Koene and Adolphs, 2015; McKee, 2013; Munson et al., 2013; Nunan and Yenicioglu, 2013; Orton-Johnson, 2010; Vayena et al., 2012). Awareness of the potential privacy implications of sharing personal information on social media is growing, driven by newsworthy cases such as Facebook's experiments in emotion manipulation (Jouhki et al., 2016) and its identification of 'vulnerable' teenagers for advertisers (Pells, 2017), or the use of social media by data analytics companies seeking insights into citizens' political attitudes and networks, to influence voter behaviour (Fromm, 2016; Arthur, 2010). In this environment, pinning down the ethical guidance for researchers is now more critical than ever, with a requirement for any guidance to be responsive and adaptable to the changes invoked by the rapid evolution of social media platforms and data science.

Most research institutions, irrespective of academic discipline, publish or adhere to some form of research ethics guidelines or standard operating procedures, as a means of ensuring the appropriate governance of studies undertaken by their staff and collaborators. While these vary in structure, content and application, they are all intended to ensure responsible and trustworthy research practice and 'to protect all groups involved in research: participants, institutions, funders and researchers throughout the lifetime of the research and into the dissemination process' (ESRC, 2010: 2). Social media research is still a relatively new and changing field and commentators have pointed to the destabilisation of traditional ethics and an unsettling of ethical expectations and assumptions for both researchers and Internet users (Whiteman, 2012). This has been compounded by a lack of relevant ethics guidance and poses particular challenges for research involving 'sensitive' data, such as information about people's health conditions, political affiliations or religious beliefs (see https://www.gov.uk/data-protection/the-data-protection-act)

Scope of ethics guidelines considered in this study

Given the growth of research using social media platforms, and its potential implications for information privacy, confidentiality and ownership, it is timely to examine the extent to which existing research ethics guidelines take such uses into account and what additions may be warranted. Social media research is taking place across multiple academic disciplines and applications for research ethics approval may thus defer to a range of different bodies. This presents challenges for the effective oversight of such research where, it has been claimed, 'no official guidance or answers regarding internet research ethics have been adopted at any national or international level' (AoIR, 2012). Mindful of the need for a cross-disciplinary perspective, we chose to study one identifiable national corpus of multidisciplinary research ethics guidelines, represented by Research Councils United Kingdom (RCUK).

Table 1. RCUK umbrella organisation and the seven UK Research Councils.

- Research Councils United Kingdom (http://www.rcuk.ac.uk)
- Arts and Humanities Research Council (http://www.ahrc.ac.uk/Pages/Home.aspx)
- Biotechnology and Biological Sciences Research Council (http://www.bbsrc.ac.uk/home/home.aspx)
- Engineering and Physical Sciences Research Council (http://www.epsrc.ac.uk)
- Economic and Social Research Council (http://www.esrc.ac.uk)
- Medical Research Council (http://www.mrc.ac.uk)
- Natural Environment Research Council (http://www.nerc.ac.uk)
- Science and Technology Facilities Council (https://www.stfc.ac.uk/home.aspx)

RCUK is a strategic partnership between the UK's seven research councils, which according to its homepage, 'has invested around £3 billion in research covering the full spectrum of academic disciplines from the medical and biological sciences to astronomy, physics, chemistry and engineering, social sciences, economics, environmental sciences and the arts and humanities' (see http://www.rcuk.ac.uk). They share an aim to 'advance knowledge and generate new ideas which lead to a productive economy, healthy society and contribute to a sustainable world'. While RCUK itself has published a set of general research ethics guidelines, each of the seven disciplinary bodies in the RCUK family of research councils (see Table 1) provides its own form of ethical advice, either through developing bespoke guidelines or deferring to other relevant guidelines in the literature. For the purposes of our study, the corpus of RCUK ethics guidelines and external guidelines recommended within these was felt to be an appropriate sample to enable a meaningful analysis of the guidance available for academic researchers in the UK.

Aims

We set out to examine how RCUK and affiliated research ethics guidelines acknowledge and deal with research involving social media overall, and specifically research involving data extracted from social media platforms (which we refer to using the generic term 'mining'). We also wanted to understand how researchers using these new forms of data in their studies are responding to the ethical challenges this presents, by examining how ethical concepts or guidelines are referred to in published research articles. We chose health research as an exemplar area, since it is highly interdisciplinary (transecting the social, medical and computational sciences, amongst others) and in which study results based on social media are being used to inform scientific knowledge and theory, public services and policies, business practices and methodological innovations (e.g. Pagliari and Vijaykumar, 2016; Tursunbayeva et al. 2017).

We are not aware of any previously published studies to have analysed the extent to which the RCUK guidelines address the use of social media data for research purposes, or how ethical concepts and guidelines are being referred to by researchers undertaking relevant projects. Our research therefore sought to answer the following two broad questions.

> *RQ1: How do RCUK ethics guidelines address the use of social media in research overall and specifically research using data harvested from social media?*
>
> *RQ2: How are ethical issues and guidelines described in published health research using social media data?*

Methods

Theoretical frameworks

To aid our analysis we drew on two ethical frameworks which, although developed in the context of social media research for health, are sufficiently generic to be applied to any field of research involving the use of social media.

The first is Bjerglund-Andersen and Söderqvist's (2012) typology of social media uses in research, which delineates five broad categories:

1. research dissemination;
2. scientific discussion and networking;
3. engaging the public;
4. academic teaching;
5. research and data collection.

For the reasons already described, we divided the last of these into two qualitatively different categories: first, using social media platforms to *enable the conduct of research*; and, second, using social media as a *source of data for research.*

Conway (2014) has gone further by suggesting a taxonomy of ethical considerations specifically relevant to the secondary use of social media data. Although this was developed in the context of Twitter mining for public health surveillance and research, it is applicable to many types of research involving data harvested from social media. This includes 10 specific considerations:

1. privacy;
2. informed consent;
3. ethical theory;
4. institutional review board (IRB)/regulation;

5. traditional research versus social media (e.g. Twitter) research;
6. geographical information;
7. researcher lurking;
8. economic value of personal information;
9. medical exceptionalism;
10. benefit of identifying socially harmful medical conditions.

While considerations 9 and 10 refer to medical issues, they can also be applied to other topics which are also uniquely sensitive (e.g. research on political attitudes) or are aimed at preventing harm (e.g. analysing extremist discourse), respectively. For the broader purposes of our study we therefore re-labelled them as 'exceptionalism' and 'benefit of identifying potential harms'.

> *RQ1. How do RCUK ethics guidelines address the use of social media in research overall and specifically research using data harvested from social media?*

To identify the corpus of ethics guidelines represented by RCUK, the websites of RCUK itself and the seven UK Research Councils were first identified via Google. The websites were then searched by entering the key words 'ethics', 'guidelines', 'funding applications' and variants of these, into their respective search boxes, and the outputs sifted manually. Searches were undertaken by the first author in February 2017.

Where a research agency was found to have more than one current ethics guideline, each of these was included, and in cases where the RCUK guidelines explicitly referred to external guidelines, the relevant source documents were also obtained for further analysis. Individual research councils were also contacted via email, asking them to state whether their organisation had developed or specifically recommended any ethics guidelines concerning the use of social media in research. Responses were received from six out of eight agencies, the non-respondents being the BBSRC and the ESRC.

The following information was extracted from each identified guideline: the name of the originating organisation, the title of the guideline, the date of the most recent version and whether the guideline explicitly referred to the use of social media or related concepts such as online or internet research.

The four guidelines referring to social media were scrutinised, to determine how they corresponded with the (adapted) typology of social media uses in research outlined by Bjerglund-Andersen and Söderqvist (2012). They were further appraised in terms of their reference to Conway's (2014) list of 10 ethical considerations for research involving social media data. The guideline search and appraisal process is summarised in Figure 1.

> *RQ2: How are ethical issues and guidelines described in published health research using social media data?*

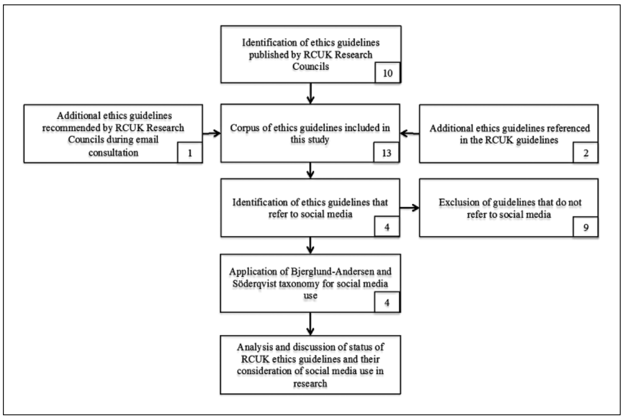


Figure 1. Summary of the guideline search and appraisal process.

For the reasons already described, we chose the example of health research using social media data, to explore how relevant ethical considerations and recommendations are being addressed in practice. The online database PubMed was searched up to 28 February 2017, using the structured query shown in Box 1.

Box 1. The search query applied to PubMed.

((“health 2.0” or “web 2.0” or “social media” or “social network” or “blog” or “wiki” or “virtual world” or “discussion forum” or “online forum” or “chat room” or “facebook” or “twitter” or “patientslikeme” or “youtube” or “instagram”) AND (“surveillance” or “infoveillance” or “mining” or “netnography” or “listening”) AND (“health” or “disease”) NOT “animal”)

The inclusion criteria encompassed peer-reviewed journal articles and conference papers describing empirical research using data from social media platforms such as Twitter or Facebook, whether extracted or studied situ, using either manual or automated methods. Studies not in English, dissertations/theses, reports or abstracts, letters to the editor and feature articles and articles intended as marketing or advertising material were excluded. No publication timeframe was applied. See Box 2 for the inclusion and exclusion criteria.

Search results were imported into the reference management software, EndNote. Abstracts and titles were initially screened for eligibility and full-text articles were obtained for those considered potentially relevant. Articles found to meet the inclusion criteria were summarised according to author name, author affiliation,

Box 2. Article inclusion and exclusion criteria.

Inclusion criteria
Types of publication: Peer-reviewed research articles. Full conference papers
Language: English
Publication timeframe: None
Types of research: Empirical studies using health-related data from social media platforms, extracted or studied *in situ*, using both manual and automated methods.
Exclusion criteria
Types of publication: Dissertations/theses; Reports or abstracts only; Letters to the editor; Marketing or advertising material; Reviews or editorials
Language: Not English
Types of research: Studies based on data from online sources other than social media (e.g. internet search histories, online news reports). Commercial research aimed at obtaining market intelligence or informing product promotion. Studies examining social media platforms, rather than using them as a source of data. Studies describing social media as a communication or broadcasting channel (e.g. for public health promotion).

publication title, publication year and abstract. Each article was also hand searched, to determine whether the authors referred to ethical considerations or guidelines when describing their study design or analysis. Where this was the case, the relevant text was extracted, tabulated and classified using Conway's taxonomy.

The components of the study, at each stage, are briefly summarised in Figure 2.

Results

> *RQ1. How do RCUK ethics guidelines address the use of social media in research overall and specifically research using data harvested from social media?*

A total of 13 separate ethics guidelines were identified, including 10 produced by RCUK itself or the individual UK research councils, 2 external guidelines recommended within these (BPS, 2012; AoIR, 2012) and one recommended by MRC during the email verification phase (INVOLVE, 2014). Of these, only four guidelines (ESRC, BPS, AoIR, NIHR) mentioned the use of social media in research.

The 13 guidelines are listed in Table 2, which also illustrates the co-referencing of guidelines within the RCUK family; for example, ARHC's guideline defers to the ESRC's guideline which, in turn, cites guidelines from the BPS and AoIR. Highlighted in bold are the four guidelines found to include guidance and recommendations specifically relating to the use of social media in research: ESRC, BPS, AoIR and BPS.

Table 3 illustrates a further level of analysis, focused on the four guidelines that encompassed social media. Based on the adapted version of Bjerglund-Andersen

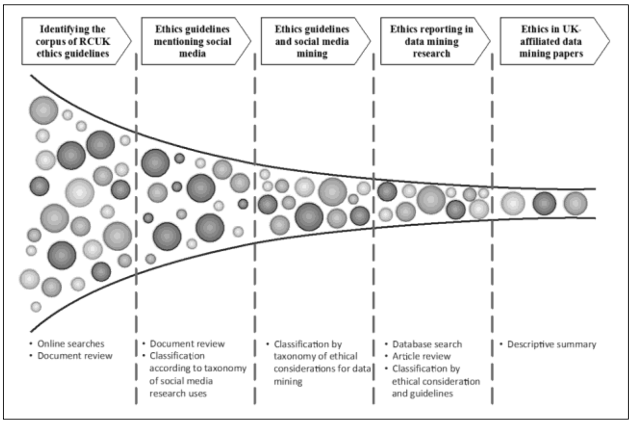


Figure 2. Focus, objectives and methods at each stage of the study.

and Söderqvist's (2012) taxonomy, all four referred to social media as a research tool, three as a source of research data, two each as a medium for scientific discussion, networking or public engagement and none for research dissemination or academic teaching. According to Conway's (2014) list of ethical considerations in social media research all four of these guidelines referred to privacy and the difference between traditional and social media research, three referred to informed consent and the use of IRBs, two referred to researcher lurking and one to ethical theory. None considered geographical information, the economic value of personal information, exceptionalism or the benefit of identifying sources of potential harm.

> *RQ2: How are ethical issues and guidelines described in published health research using social media data?*

The structured search of PubMed yielded 469 potentially relevant studies, of which 156 remained after screening against the inclusion and exclusion criteria. These studies had a variety of aims, including assessing public reactions to health reforms, identifying health behaviours such as medication compliance, understanding health attitudes and sentiments, undertaking post-market surveillance, exploring social networks relevant to health, searching for indicators of infectious and non-communicable disease trends and comparing the value of different social media platforms or tools for analysing health-related events or patterns. Only 50 articles referred to one or more of the ethical concepts, procedures or approval

Table 2. Ethics guidelines screened for references to social media uses in research.

Research Council (date)	Guideline title	Includes	Refers to social media
RCUK (2013)	Policy and Guidelines on Governance of Good Research Conduct		
AHRC (2016)	Research Funding Guide	RCUK ESRC	
BBSRC (2017)	BBSRC Research Grants The Guide		
EPSRC (2013)	Framework for Responsible Innovation	RCUK	
MRC (2012b)	Policy and Guidance on Sharing of Research Data from Population and Patient Studies		
MRC (2000)	Personal Information in Medical Research		
MRC (2012a)	Good research practice		
NERC (2015)	Ethics Policy		
STFC (2013)	Public Engagement with Science and Technology	MRC RCUK	
ESRC (2015)	**Framework for research ethics**	BPS AoIR RCUK	**x**
BPS (2012) ***In ESRC***	**Guidelines for ethical practice in psychological research online**		**x**
AoIR (2012) ***In ESRC***	**Ethical Decision Making and Internet Research**		**x**
NIHR (2014) ***Recomm MRC***	**Guidance on the use of social media to actively involve people in research**		**x**
TOTAL			**4**

processes specified in Conway's taxonomy (Figure 3). However, while most of these mentioned IRB approval, only 13 referred to other relevant ethical considerations and five of the ethical considerations in Conway's taxonomy were not mentioned at all. In order of frequency, the breakdown of ethical considerations was as follows: Research Ethics IRB Approval/Regulation (43), Privacy (26), Informed Consent (16), Ethical Theory (7), Traditional Research vs Social Media Research (3), Researcher Lurking (3), Identifying Potential Harms (2), Geographical Information (0), Economic Value of Personal Information and (0) Medical Exceptionalism (0).

Nine of the studies we identified using PubMed were affiliated with UK-based organisations and their consideration of ethical concepts are further described in Table 4. In short, they described:

Table 3. Types of research use and ethical considerations for data re-use.

Research council	Guideline title (date)	Types of social media use in research (RQ1)						Ethical considerations for the use of social media data in research (RQ2)									
		1	2	3	4	5	6	A	B	C	D	E	F	G	H	I	J
ESRC	Framework for research ethics (ESRC, 2015)					x	x	x	x	x	x	x					
BPS (In ESRC)	Guidelines for ethical practice in psychological research online (BPS, 2012)			x		x	x	x	x			x		x			
AoIR *(In ESRC)*	Ethical Decision Making and Internet Research (AoIR, 2012)		x			x	x	x	x		x	x		x			
NIHR *(Recomm. MRC)*	Guidance on the use of social media to actively involve people in research (INVOLVE, 2014)		x	x		x		x			x	x					
Total		0	2	2	0	4	3	4	3	1	3	4	0	2	0	0	0

Bjerglund-Andersen and Söderqvist's classes of social media use (adapted): 1, Research Dissemination; 2, Scientific discussion/networking; 3, Engaging the public; 4, Academic teaching; 5, Social media as a research tool; 6, Social media as a source of research data.
Conway's ethical considerations for social media data use: A, Privacy; B, Informed consent; C, Ethical theory; D, IRB approval/regulations; E, Traditional vs social media research; F, Geographical information; G, Research lurking; H, Economic value of personal information; I, Exceptionalism; J, Benefit of identifying sources of potential harm.

1. a study using data from Facebook and Twitter to examine the usefulness of social media for post-market drug safety surveillance (Powell et al., 2016),
2. a content analysis of social media data posted on two web forums to monitor the misuse and non-medical use of the antidepressant and smoking cessation drug bupropion (Anderson et al., 2017),
3. an analysis of the online response to a case of a breastfeeding mother being ejected from a UK retail premises (Grant et al. 2016),
4. a case study into the difficulties, challenges and rewards of using social media by student nurses through analysing data from a Twitter chat (Sinclair et al., 2015),
5. a netnographic study of user decision-making, home preparation and consumptive patterns of laudanum (Van Hout and Hearne, 2015),
6. a study investigating the feasibility of developing predictive models that identify potential superusers of online healthcare support groups (van Mierlo et al., 2017)

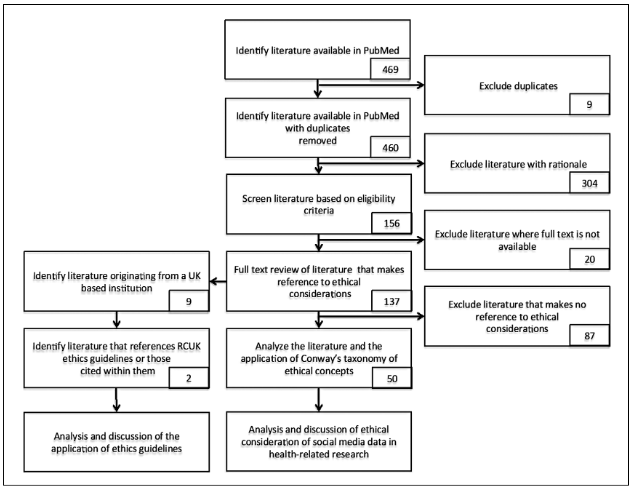


Figure 3. Number of studies included at each stage of the screening process.

7. a qualitative study into how young people used a youth-orientated, moderated, online, eating disorders discussion forum, run by an eating disorders charity (Kendal et al., 2017),
8. a thematic analysis of readers' comments to UK online news reports on the acceptability of financial incentives for breastfeeding (Giles et al., 2015),
9. a qualitative and quantitative summary of online reaction to media reports to the UK government strategy on childhood obesity in England (Gregg et al., 2017).

The first two of these were authored by researchers from the pharmaceutical sector while the remaining seven were from UK universities. Of these nine UK studies, two (Giles et al., 2015; Gregg et al., 2017) referenced the research ethics guidelines produced by the BPS, as identified in part 1 of our study, while none referred to the guidelines developed by RCUK or its member councils.

The number of papers identified at each stage of the search process is shown in Figure 3, while Table 4 provides a breakdown of the ethical considerations represented in each of the included articles, along with illustrative quotations.

Table 4. Ethical guidelines and concepts referred to in studies using social media data for health-related research found in PubMed.

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
1	Campaigns and counter campaigns: reactions on Twitter to e-cigarette education (Allem et al., 2016)	x			x							'Ethics approval The University of Southern California Institutional Review Board approved all procedures.' (p. 229)
2	Using Social Listening Data to Monitor Misuse and Nonmedical Use of Bupropion: A Content Analysis (Anderson et al., 2017)	x	x	x	x							'…Content that is deemed sensitive and is in the public domain sits in a gray zone from an ethical perspective, and the extent of protection for the individuals who write the content and the communities that host the content should be assessed on a case-by-case basis. The community discussions demonstrate that contributors are aware of the public nature of the content that they post, and almost all contributors utilize pseudonyms to mask their identities. Although the subject matter may be seen as sensitive, these elements led the research authors to determine that consent from individual contributors was not necessary to conduct the research. It was also important to maintain any particular contributor's anonymity, as the extent to which their pseudonym may reveal identifying information about them is unknown to the researchers…. Because our research did not involve intervention or interaction with the individuals, nor is the information individually identifiable, our study did not meet the criteria of the Office for Human Research Protections (OHRP) framework that guides institutional review board (IRB) status. As such, IRB approval was not pursued. Some researchers anonymize the names of the Web forums that they utilize as data in order to further assure confidentiality of the individual contributors or because the group had neither been actively involved in the research nor given consent to be involved …We reviewed the site's privacy notice and user agreement and determined that gathering data for research purposes was within the scope of permitted uses…. We contacted a third potential data source, Erowid, to request consent and terms of access for gathering samples from their database of user-reported experiences with drugs…. their usage agreement explicitly prohibited data gathering or publishing of analyses without prior permission…Erowid was excluded as a data source for this study.' (p. 5)
3	Characterizing the followers and tweets of a marijuana-focused Twitter handle (Cavazos-Rehg et al., 2014)				x							'The Twitter data in the current study is public. The Washington University Institutional Review Board reviewed our study protocol and our research was deemed exempt from human subjects review.' (p. 3)
4	What Online Communities Can Tell Us About Electronic Cigarettes and Hookah Use: A Study Using Text Mining and Visualization Techniques (Chen et al., 2015)	x			x							'Research Ethics Statement: Publicly available social media content can be an invaluable complement to data provided by study participants in more explicit research contexts because it is a rich source of information on how behaviors with health impacts may naturally occur in the real world. In order to protect the identities of forum users, we have not provided explicit quotations, but instead described the content in as much detail as possible, both quantitatively and qualitatively, in line with ethical guidelines [44,45]. The work reported in this paper has been certified as exempt from review under 45 CFR 46.101(b), category 4 by the University of California San Diego Institutional Review board (Project #140844X).' (p. 5)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
5	Impact of Twitter intensity, time, and location on message lapse of bluebird's pursuit of fleas in Madagascar (Da'ar et al., 2016)				x						x	'Ethical approval: Not required.' (p. 6) (N.B. The study concerns a bubonic plague outbreak.)
6	When 'Bad' is 'Good': Identifying Personal Communication and Sentiment in Drug-Related Tweets (Daniulaityte et al., 2016)	x			x							'The Wright State University institutional review board reviewed the protocol and determined that the study meets the criteria for Human Subjects Research exemption 4 because it is limited to publicly available tweets. Tweets used as examples were modified slightly to ensure the anonymity of Twitter users who had posted them.' (p. 3)
7	Surveillance Tools Emerging From Search Engines and Social Media Data for Determining Eye Disease Patterns (Deiner et al., 2016)				x							'With approval from the University of California San Francisco (UCSF) Institutional Review Board, we obtained total weekly counts of all encounters with diagnosis names containing the string "conjunctivi"' (p. 1025)
8	How to exploit twitter for public health monitoring? (Denecke et al., 2013)			x								'we did not yet consider the legal and ethical issues related to the use of data from TV/ radio and social-media for public health surveillance as well as the reliability of the collected information' (p. 339)
9	Computer-assisted update of a consumer health vocabulary through mining of social network data (Doing-Harris and Zeng-Treitler, 2011)				x							'These records were obtained by another group in our department with internal review board (IRB) approval. IRB approval was given for a member of that group to compare terms to this database for us, returning a yes/no answer.' (p. 4)
10	#discrimination: The Online Response to a Case of a Breastfeeding Mother Being Ejected from a UK Retail Premises (Grant, 2016)				x							'As the data were hosted on a website intended for public consumption that did not require membership to access content, ethical approval was not This position was con- firmed by the chair of the Cardiff University School of Medicine Research Ethics Committee.' (p. 143)
11	Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks (Gu et al., 2014)	x			x							'All the information obtained online was in simplified Chinese language and released publicly by the websites, but no personal identification information, such as name or email address, was collected. This study was approved by the Institutional Review Board in the Zhejiang Provincial Centers for Disease Control and Prevention.' (p. 3)

(Continued)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
12	Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model (Guan et al., 2015)		x		x							'Ethical considerations of the study have been reviewed and granted by the Review Board of the Institute of Psychology, Chinese Academy of Sciences' (p. 2)
13	Public health surveillance of dental pain via Twitter (Heaivilin et al., 2011)	x										'Username removed to maintain privacy of Twitter user.' (p. 1049)
14	A cross-sectional examination of marketing of electronic cigarettes on Twitter (Huang et al., 2014)		x		x							'Ethics approval This study is cleared for ethics by Research Ethics Boards or International Review Boards at the University of Illinois at Chicago (USA)' (p. 30)
15	Text classification for assisting moderators in online health communities (Huh et al., 2013)				x							'Thread conversations in WebMD online communities are publicly available. We applied for approval from the University of Washington's Institutional Review Board (IRB) and received a letter stating that our project is unregulated by IRB because the data is publicly available.' (p. 999)
16	The Use of Social Media by State Health Departments in the US: Analyzing Health Communication Through Facebook (Jha et al., 2016)				x							'Compliance with Ethical Standards' (p. 178)
17	Cadec: A corpus of adverse drug event annotations (Karimi et al., 2015)				x							'Ethics approval for this project was obtained from the CSIRO ethics committee which classified the work as low risk' (p. 80)
18	Blogging for weight loss: personal accountability, writing selves, and the weight-loss blogosphere (Leggatt-Cook and Chamberlain, 2012)	x	x	x	x							'Researching in cyberspace presents challenges. One concerns ethical practice in using material posted in public forums, where a major contested concern is the boundary between public and private, and related issues of consent. In our use of blog content, we concurred…that archived blog content, unprotected by subscription or password access, is public, is intended to be so by its authors, and so does not require consent… In reporting blog content, we exercised judgement to ensure that bloggers were represented sensitively. Reporting also raises ethical concerns around verbatim citation of blog content, since this can facilitate searches to identify individuals …we considered that the overtly public content … did not warrant such protection and, moreover, that blog writers require credit for their published work online. These various arguments were used in gaining institutional ethical consent for the research.' (p. 965)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
19	Using Real-Time Social Media Technologies to Monitor Levels of Perceived Stress and Emotional State in College Students: A Web-Based Questionnaire Study (Liu et al., 2017)		x		x							'Ethics approval was obtained from the UCLA Research Ethics Board.' (p. 2)
20	Participatory surveillance of diabetes device safety: a social media-based complement to traditional FDA reporting (Mandl et al., 2014)		x		x							'Online consent for use of member-provided data for research is obtained when a member first accesses the app.' (p. 688) 'Ethics approval Boston Children's Hospital Institutional Review Board.' (p. 691)
21	Symptom clusters in women with breast cancer: an analysis of data from social media and a research study (Marshall et al., 2016)		x	x	x							'Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the research study of breast cancer survivors. Permission to extract data from the forum MedHelp.com was granted by site managers, and all data from this source were collected anonymously.' (p. 547)
22	Reaction on Twitter to a Cluster of Perinatal Deaths: A Mixed Method Study (Meaney et al., 2016)	x			x							'Only data which were publically [*sic*] available were collected and no attempts were made to contact any individual; therefore, no ethical approval was sought for this study. Despite these data being available to public, there is still an onus to ensure that ethical standards are met. Therefore, in line with other similar studies … personal identity information, including individuals' Twitter usernames, have been removed from the example tweets presented below.' (p. 3)
23	What can we learn about the Ebola outbreak from tweets? (Odlum and Yoon, 2015)	x			x						x	'Ethical approval This study used publically [*sic*] available data, and analyses meet the criterion for exemption §46.101(b)4 Research, involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available, or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.' (p. 563)
24	Towards early discovery of salient health threats: A social media emotion classification technique (Ofoghi et al., 2016)	x										'All URLs, email addresses, mentions (i.e., @replies and @usernames), and hash tags were replaced by "url", "emailAddress", "atSign", and "hashTag", respectively.' (p. 507)

(Continued)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
25	Discovering health topics in social media using topic models (Paul and Dredze, 2014)				x							'Ethics Statement: The work described in this paper was reviewed by the Homewood Institutional Review Board at Johns Hopkins University and received an exemption since all data is publicly available.' (p. 2)
26	Social Media Listening for Routine Post-Marketing Safety Surveillance (Powell et al., 2016)	x			x							'Much of the data posted by these patients are publicly available on the Internet, depending on the individual's use of privacy settings when posting.' (p. 444) 'For the purpose of this research project, the following additional steps were taken to protect privacy: • Once the data had been de-identified by the vendor, no attempt was made to re-identify the person making the post. As a result, no attempt was made to obtain follow- up information about potential AEs. • Posts from the same person were not linked.' (p. 446)
27	YouTube: a promotional vehicle for little cigars and cigarillos? (Richardson and Vallone, 2014)				x							'This study did not require institutional review board approval given that it used freely accessible media.' (p. 2)
28	SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks (Rodrigues et al., 2016)	x			x							'This work is part of a project evaluated and approved by the Ethics Committee under the number 31191214.7.0000.5083/UFG. Moreover, all data collected from online social network were published by the user as a public text in a public group.' (p. 82)
29	Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from Twitter (Sarker et al., 2016)	x			x							'Ethical approval: Not applicable Informed consent: Not applicable' (p. 239)
30	Supplementing Public Health Inspection via Social Media (Schomberg et al., 2016)				x							The University of California Irvine Internal Review Board granted non-human subjects exemption to this study. This classification exempted this study from further University of California Irvine Internal Review Board review.' (p. 4)
31	Tanning bed burns reported on Twitter: over 15,000 in 2013 (Seidenberg et al., 2016)				x							'Ethics approval: This article does not contain any studies with human participants performed by any of the authors.' (p. 275)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
32	Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words Among Korean Adolescents (Song et al., 2016)	x			x							'The Institutional Review Board of the authors' institutions approved the study protocol. No personally identifiable information was available in the collected data.' (p. 670)
33	Social Big Data Analysis of Information Spread and Perceived Infection Risk During the 2015 Middle East Respiratory Syndrome Outbreak in South Korea (Song et al., 2017)				x							'The Institutional Review Board at the authors' institution approved the study protocol of this study.' (p. 27)
34	Sources of information and behavioral patterns in online health forums: observational study (Sudau et al., 2014)	x			x							'The Ethics Committee of the University Medical Center Göttingen confirmed (ref 11/5/13) that ethical approval was not necessary due to the nature of the data (secondary data analysis of anonymized data).' (p. 12)
35	Do cancer patients tweet? Examining the twitter use of cancer patients in Japan (Tsuya et al., 2014)				x							'This study was approved by the Institutional Review Board at Yamagata University Faculty of Medicine (H24-133).' (p. 2)
36	Sharing data for public health research by members of an international online diabetes social network (Weitzman et al., 2011a)	x	x		x							'All study activities were reviewed and approved by the Children's Hospital Boston Committee on Clinical Investigation under a model of implied consent that was based on the pre-existing norms for sharing in the community and in alignment with the published privacy policy and terms of use of the site that clearly inform the community about conditions for sharing data and privacy protections.' (p.4)
37	Surveillance of an Online Social Network to Assess Population-level Diabetes Health Status and Healthcare Quality (Weitzman et al., 2011b)		x		x							'Objective: Test a novel health monitoring approach by engaging an international online diabetes social network (SN) in consented health surveillance.' (p. 1) 'Study activities were reviewed and approved by the Children's Hospital Boston Institutional Review Board.' (p. 2)
38	Participatory surveillance of hypoglycemia and harms in an online social network (Weitzman et al., 2013)		x		x							'Consent is obtained within the app for use of member data. Study activities were approved by the Boston Children's Hospital institutional review board' (p. 346)

(Continued)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
39	Public Trauma after the Sewol Ferry Disaster: The Role of Social Media in Understanding the Public Mood (Woo et al., 2015)	x										'There is no information that could potentially reveal the identity of social media user, namely user confidentiality is maintained.' (p. 10977)
40	'Trip-Sitting' in the Black Hole: A Netnographic Study of Dissociation and Indigenous Harm Reduction (Hearne and Van Hout, 2016)	x				x		x				'Data collection and analysis were regarded as observations of online public behavior, in accordance with the ethical protocols and recommendations of SACHRP (Secretary's Advisory Committee for Human Research Protections 2013). As per the ethical and methodological protocols of passive netnography, the researcher did not make contact with fora users and upheld observational status at all times. Anonymity of the fora and its members was safeguarded by removal of user pseudonyms and URLs to the site, and illustrative quotations were paraphrased by the team with consensus in interpretation to avoid backtracking of these through Internet search engines' (p. 234–235)
41*	Demographic and Indication-Specific Characteristics Have Limited Association With Social Network Engagement: Evidence From 24,954 Members of Four Health Care Support Groups (van Mierlo et al., 2017)	x			x							'All data collection policies and procedures adhered to international privacy guidelines and were in accordance with the Helsinki Declaration of 1975, as revised in 2008. The study was consistent with the University Research Ethics Committee procedures at Henley Business School, University of Reading, and was exempt from full review.' (p. 5)
42*	Public reaction to the UK government strategy on childhood obesity in England: A qualitative and quantitative summary of online reaction to media reports (Gregg et al., 2017)	x	x	x	x	x						'Ethical approval was awarded by the Manchester Metropolitan University ethics committee prior to data collection. We have followed the guidance for internet- mediated research from the British Psychological Society and adhered to copyright laws in conducting this work. Direct consent could not be attained because of the nature of the data collection however; implicit consent was deemed to have been given by virtue of posting in an open forum. No directly identifiable data were collected, comments were disassociated from usernames prior to analyses.' (p. 451)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
												'The ethics of methods that make use of readily available online data are less developed than an "off line" project which relies on the informed consent of its participants for its moral sufficiency. The participants in this study could not have anticipated that their contributions would be used for this research therefore; they have not provided their consent in the usual sense. However, participants have contributed to public online fora with the intention of having their views heard and by doing so they have sought to influence the debate surrounding the government's childhood obesity strategy. Notwithstanding that, the participants have not provided their express consent, in our view and as agreed by the ethics review board, it is reasonable to infer that by their actions they implied consent to the use of their contributions in subsequent debate. While data collection without prior notice raises ethical issues, it has the advantage of allowing a level of candour that may not be forthcoming where a researcher declares their presence and seeks explicit consent. Some of the comments were inappropriate but have the advantage that they were open and reflect the reality of citizens' experience. It is conceivable that we may have taken steps to notify participants of our intention to use material for research. For example by seeking to amend the terms of the forum with the cooperation of the site owners, or simply by joining the discussions and announcing our presence and thereby imputing acquiescence from continued participation rather than express consent. On balance, we took the view that such action might influence discussions and undermine the authenticity of the debate and, was disproportionate to the risk of any harm.' (pp. 456–457)
43	Doing Recovery Online (Mudry and Strong, 2013)	x	x		x							'We obtained ethical approval from our university research ethics board, and the first author gained informed consent from the participants involved. Participation was voluntary, anonymous, and confidential. We did not collect any personally identifying information and all participants remained anonymous. We extracted and present verbatim quotes (punctuation corrected, grammar uncorrected) as exemplars.' (p. 315)
44	Patterns of Treatment Switching in Multiple Sclerosis Therapies in US Patients Active on Social Media: Application of Social Media Content Analysis to Health Outcomes Research (Risson et al., 2016)	x			x							'… The secondary data source used for the analysis meets all the US Health Insurance Portability and Accountability Act (HIPAA) compliance standards, ensuring patient anonymity. As such, approval from an Institutional Review Board was not necessary.' (p. 3)

(Continued)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
45*	How a moderated online discussion forum facilitates support for young people with eating disorders (Kendal et al., 2017)	x	x		x							'Online communities for young people are recognized as a valid focus for research, and the forum was in the public domain. However, there is also a debate on the ethical issues around using publicly accessible online discussion for research. The identities of the forum users were unknown so we were not in a position to obtain individual consent from them. Instead, the charity gave us proxy consent to access and use the posts. To enhance transparency, we advertised and explained our research on the charity's website and Twitter feed before commencing the study. We protected the privacy of the forum users by removing terms and phrases that could identify them, including the name of the charity. We obtained ethical approval for this study in March 2012 from the UK NHS National Research Ethics Service and the University of Manchester Research Ethics Committee.' (p. 102)
46*	To Twitter to Woo: Harnessing the power of social media (SoMe) in nurse education to enhance the student's experience (Sinclair et al., 2015)	x	x	x								'As all Twitter chat transcripts are made freely available to anyone as a record afterwards at www.wenurses.co.uk. it wasn't felt there were any consent issues involved for the participants. Indeed, the case study was constructed around an analysis of the participant's tweets, all of which are in the public domain and open to scrutiny.' (p. 509) 'The NMC (2012) state that it should be assumed everything posted online is public, permanent and shared, even with the strictest privacy settings and we also needed to get this message across to students effectively. Understanding this can prevent unintentional postings that may cross professional boundaries. The "privacy illusion" is discussed by Aylott (2011) suggesting that nurses often believe that their strict privacy settings will protect their posts, when in fact this is purely an illusion….' (p. 511)
47	Analysis of Patient Narratives in Disease Blogs on the Internet: An Exploratory Study of Social Pharmacovigilance (Matsuda et al., 2017)		x		x							'The study protocol was reviewed and approved by the nonprofit MINS Institutional Review Board. The board waived informed consent because the data source did not contain personal information. In addition, we presented the data at the group level rather than at the individual level.' (p. 6)
48	Health Risk Information Engagement and Amplification on Social Media: News About an Emerging Pandemic on Facebook (Strekalova, 2017)				x							'An institutional review board approval was obtained to ensure that data collection and analysis were compliant with ethical standards for behavioral research.' (p. 333)

Table 4. (Continued)

#	Article title	Categories of ethical consideration, from Conway's taxonomy										Relevant text
		A	B	C	D	E	F	G	H	I	J	
49*	Acceptability of financial incentives for breastfeeding: thematic analysis of readers' comments to UK online news reports (Giles et al., 2015)	x	x	x	x	x		x				'The chair of the Newcastle University Faculty of Medical Sciences research ethics committee confirmed that ethical approval was not required for this study. However, we did consider, in detail, numerous ethical issues arising from this research and report on these in the discussion section.' (p. 5) 'Netnography is still a relatively new approach to data collection and analysis and there is limited guidance on the ethics of using this approach. As such, we feel it is appropriate to explicitly discuss the main ethical issues raised in some detail. The comments analysed here were not provided for research purposes, and commenters are not aware that we have used them for this. They have, therefore, not provided informed consent to take part in the research…We sought permission from the websites involved to use their content and adhered to copyright guidelines throughout…We did not identify ourselves as researchers and observers to the online communities. This was primarily because we chose, as others have done, not to interfere with comments and discussions as they developed. This meant that as researchers we did not influence the data included in the research – as might have been the case in more traditional interviews or focus groups…To preserve the anonymity of commenters, we have been careful not to include any details in quotations that could have identified the commenter. We also followed best practice guidance provided by the British Psychological Society.' (pp. 11–12)
50*	'Vintage Meds': A Netnographic Study of User Decision-Making, Home Preparation, and Consumptive Patterns of Laudanum (Van Hout and Hearne, 2015)	x						x				'In compliance with unobtrusive and naturalistic features of netnographic research, researchers acknowledged the dynamics of the public online "drug user" environment, maintained observational status, and respected the inherent flexibility and openness of the approach. Confidentiality measures applied to the dataset included storage in an online, password-protected computer and removal of screen pseudonyms, URLs, country and city identifiers' (p. 600)
	Total	**26**	**16**	**7**	**43**	**3**	**0**	**3**	**0**	**0**	**2**	

Key to Conway's categories of ethical consideration: A, Privacy; B, Informed consent; C, Ethical theory; D, IRB/regulation; E, Traditional research vs social media research; F, Geographical information; G, Researcher lurking; H, Economic value of personal information; I, Medical exceptionalism; J, Benefit of identifying socially harmful medical conditions.

Discussion

Our analysis indicates significant gaps in the ethical governance of research using data mined from social media, illustrated by the incompleteness and inconsistency of current guidelines and an absence of ethical discourse in published research articles.

Status of RCUK ethics guidelines on social media and social media data

Of the seven multi-disciplinary ethics guidelines published by RCUK, only one (ESRC) specifically considered the use of social media in research, despite such research now straddling the remits of many national funding agencies. Two research councils (ESRC, MRC) nevertheless recommended guidelines from other bodies (AoIR, BPS, NIHR/INVOLVE), generating a corpus of four social-media relevant guidelines for UK researchers. These referred to social media as a research tool (4/4), as a source of data (3/4), as a means of public engagement (2/4) and as a channel for scientific discussion and networking (2/4), but did not mention their use for research dissemination or teaching, which also appear in our adapted version of Bjerglund-Andersen and Söderqvist's (2012) taxonomy. With specific reference to the mining and re-use of social media data, these guidelines prioritised privacy (4/4), differences between digital and conventional research (4/4), informed consent (3/4), IRB approval/regulation (3/4) and researcher lurking (2/4), although none of the other four ethical considerations in Conway's (2014) framework were covered. Although MRC was the source of three research ethics guidelines, none referred to the use of social media, in contrast to their detailed consideration of ethical issues surrounding the re-use of institutional and research datasets, where most of the UK's 'big data' investments are taking place. ESRC provided the most comprehensive overview of social media ethics, also deferring to the external AoIR and BPS guidelines, likely reflecting the importance of digital social research within ESRC's portfolio. While these differences between research councils are to some extent understandable, they indicate a segmentation of data ethics along disciplinary lines, which is unhelpful in an environment where interdisciplinary projects are the norm, rather than the exception, underscoring the need for collaboration and agreement on universal principles.

Our focused analysis of articles indexed in PubMed also indicates a widespread neglect of ethical issues amongst research practitioners using social media data in health-related studies. Where ethical issues were discussed, this tended to centre on the *procedures and requirements* necessary to obtain IRB approval, such as demonstrating an awareness of privacy risks and determining whether consent was necessary, rather than showing a deeper concern with the *moral or societal* implications of repurposing information that people have shared for reasons other than research. Indeed, many published studies either did not mention ethical issues at all or simply stated that the data were available in the public domain and consent was therefore not required. While articles containing more comprehensive and thoughtful ethical discussion were found (Anderson et al., 2017; Gregg et al., 2017; Leggatt-Cook and Chamberlain, 2012), few studies using social media data considered the full range of ethical issues articulated in Conway's taxonomy. These 50 studies prioritised IRB approval/regulation (43), privacy (26), informed consent (16), ethical theory (7), traditional vs social media research (3), researcher lurking (3) and the benefit of identifying potential harms (2). None of the other three considerations in Conway's framework of ethical considerations were covered. Significantly, of the nine eligible articles originating from UK institutions, only two referred to the RCUK guidelines, suggesting either a lack of awareness or a strategic neglect, both of which indicate the need for better communication and training.

As already noted, ESRC was the only UK research council whose own ethics guidelines explicitly considered the use of social media in research. Their *Framework for Ethics* (2015) includes a detailed overview of relevant issues, along with examples, and illustrates the potential for ethics guidelines to evolve in response to emerging innovations. While the earlier version of this framework (ESRC, 2010) advised that research involving respondents through the Internet, may 'involve more than minimal risk', no specific examples of risk were provided to guide researchers in this assessment. This lack of specific guidance was also reflected in the 'frequently asked questions' section dealing with Internet searches, where it was simply noted that the rapidly evolving nature of the field and the use of web pages and instant messaging for research purposes 'pose new ethical dilemmas' that need to be addressed. In contrast, the guidelines published in January 2015 refer explicitly to ethical considerations associated with the use of social media as a research tool and as a source of research data. These include uncertainties over how to apply ethical concepts such as 'privacy' and 'anonymity', which may be interpreted differently by social media users and researchers, and the potential sensitivity of topics discussed in these settings, such as health issues. They caution that, while information intentionally published on the Internet is 'in the public domain', the identity of individuals should be protected unless it is critical to the research, such as in studies analysing statements by public officials. ESRC's 2015 guidelines also advise researchers to abide by the regulations and permissions set by the data holders (e.g. Twitter, Facebook), particularly when these are required for compliance with data protection legislation, bearing in mind that such research may cross legislative jurisdictions. The framework also benefits from deferring to two internet-specific research ethics guidelines developed by the BPS and the AoIR.

While the BPS guidelines do not explicitly refer to social media, this is implied in the term ‘internet-mediated research’, which encompasses the use of online platforms as means of engaging the public, as a set of research tools and as a source of data for secondary uses, consistent with our taxonomy. A total of 10 ethical considerations are highlighted, which overlap with but are somewhat different from those provided by Conway. These include verifying identity, private versus public space, informed consent, levels of control, withdrawal, debriefing, deception, monitoring, protection of participants and researchers, and data protection. These are grouped into four sectors of a grid, relating to whether participants are actively recruited or are unaware of their involvement in the study, as well as whether they are identified or anonymous. Although the BPS guidelines go some way towards providing actionable recommendations for researchers, they should not be considered exhaustive, given that only four of the 10 ethical concepts identified by Conway (privacy, informed consent, IRBs and researcher lurking) are addressed. A newer BPS guideline, currently under beta-testing, has extended the 2012 framework but, as yet, does not refer to social media specifically (BPS, 2017). Based on our study, we recommend including this.

The AoIR is a widely recognised international academic association dedicated to the advancement of the cross-disciplinary field of Internet studies. The AoIR ethics guideline referred to by the ESRC (AoIR, 2012), outlines several high-level themes, including the difficulty of understanding whether such research involves ‘human subjects’ for the purposes of ethics approval, differentiating ‘public from private’, conceptualising data or text as an extension of ‘persons’, and reconciling ‘top down versus bottom-up approaches’ for managing potential harms and benefits of research. The document includes an extensive list of considerations, such as understanding the context of the research, the primary objective of the research, how the data will be accessed, stored and disseminated, and the rights of participants, who may be unaware that their data are being used. Unlike the BPS guideline, the AoIR guideline explicitly mentions social media, and gives examples of social media data uses that present ethical challenges.

Given the potential sensitivity of medical information available online, it is somewhat surprising that the MRC does not provide specific guidance for researchers conducting studies using social media data. Nevertheless, in their email verifying this, the MRC recommended that we review the guidance provided by the NIHR as part of the INVOLVE advisory group. INVOLVE was established by NIHR in 1996 to support active public involvement in NHS, public health and social care research. In 2014, they published ethics guidelines on using social media to engage citizens in public debate and research, as a forum for scientific discussion and networking, and as tool for undertaking research and consultation. They list the types of social media platforms available, provide case studies of their use, outline the benefits and challenges, consider how to manage risk, and offer tips based on researcher experience. Applying Conway's taxonomy to the NIHR guidance, however, indicates that only three of the 10 ethical concepts are addressed, namely; privacy, the use of IRBs and the difference between traditional and social media based research. These reflect the public-engagement remit of INVOLVE, which may explain why the secondary use of social media data for research is not discussed explicitly.

The absence of any reference to research using social media in the remaining RCUK guidelines is noteworthy. Whilst in some cases this is entirely understandable, for example the STFC focuses primarily on particle and nuclear physics and science infrastructure, in others it would seem appropriate to include these new forms of data. For example, one EPSRC project in which the second author is involved specifically focuses on the use of social media, crowdsourcing and citizen science, albeit driven by computer scientists (SOCIAM; see http://sociam.org). This project includes themes in health and social science, illustrating how social media research transects disciplinary boundaries and may potentially fall within the scope of several ethics bodies.

The following quotation from the AoIR (2012) guideline neatly illustrates the need for this trans-disciplinary thinking.

> *'Manipulation and close study of information generated by social media networks certainly constitutes a different research environment than sticking a needle into a volunteering person in a medical laboratory. On the other hand, entire communities have felt harm from use of their DNA data more than a decade after it was collected and anonymously aggregated'* (AoIR, 2012: p. 13)

Ethical maturity of health research using social media data

The paucity of ethical considerations in the health-related research identified via PubMed is noteworthy; indeed, very few relevant studies went further than acknowledging consultation with their IRB, which is primarily undertaken for instrumental reasons. Those that did originated predominantly from the sub-field of primary care research or from researchers based in pharmaceutical companies routinely subjected to ethical oversight. Although very few studies were affiliated with UK research organisations, it is troubling to see that only two of the nine we identified referred to the RCUK or associated ethics guidelines.

The dominance of instrumental over moral considerations seen in the scientific papers we reviewed, suggests that researchers using these methods are heavily dependent on IRBs and journal editors to play the role of their ethical conscience. It is therefore essential that ethics committees and editors evaluating research using social media data are aware of the range of platforms available and how they work, and can draw on the latest interdisciplinary guidelines to inform their decision-making. We recommend that editors and peer-reviewers seek authors' explanations of the ethical challenges they faced and how these were managed during the conduct of their studies, therefore enabling greater transparency and encouraging knowledge sharing within the research community.

Policy implications

Despite their use now being common, the emergence of social media and other online platforms has taken traditionally slow-moving governments and academic institutions somewhat off-guard. Uncertainties about what is appropriate, acceptable, legal and responsible in these new virtual spaces, and for different forms of digital personal information, has also fuelled broader debates. These include debates around the need for 'net neutrality' or equal access to internet content and services amongst all users (McKee, 2011), how to maintain control of key Internet domain names in the global public interest (Mackey et al., 2014) and calls for a 'Magna Carta for Data' (Kiss, 2014; O'Sullivan, 2017). Moreover, it is contributing to the dilemma of governments seeking to generate economic, scientific and societal value from existing data assets whilst also protecting citizens from unwanted surveillance and intrusion. Health research is one area in which this discussion has been particularly acute, due to the traditionally stringent ethical demands placed on the protection of confidentiality. In the UK, the growing use of health records for research (Knapton, 2014), coupled with public disquiet over controversial programmes such as Care.Data (Boseley, 2016) and Google DeepMind's Streams project (Wakefield, 2017) have focused considerable policy attention on the need for ethical and robust governance when it comes to the use of patient information (e.g. Richards et al., 2015; National Data Guardian, 2017). In this context, it is noteworthy that, by comparison, the ethics of using social media data in health research has been somewhat neglected, albeit such data is seldom managed by the state or by healthcare institutions with a duty to protect it. It is nevertheless arguable that the same principles of respect, confidentiality and protection from harm or embarrassment should be followed as would be expected in any other form of bona fide research.

Caveats and opportunities for further research and development

Our review of ethics guidelines was limited to those provided or recommended by RCUK and its seven UK Research Councils and we are aware of other relevant guidelines developed by UK-based researchers (Convery and Cox, 2012) and organisations beyond the scope of this study (e.g. NCCPE; see http://www.publicengagement.ac.uk/work-with-us/completed-projects/ethics-cbpr/resources/ethical-guidelines-web-resources). We recommend further research involving a wider corpus of research ethics guidelines, to test the generalisability of our results in the

UK, and as a means of catalysing the development of internationally applicable ethics guidelines for research involving social media platforms and data.

The variable coherence, consistency and navigability of the RCUK websites presented a challenge for identifying relevant ethics guidelines, particularly in the case of MRC and EPSRC. For MRC, this was mainly due to its diverse portfolio of specialised guidelines, covering topics from clinical trial management through to the use of human tissue samples. For EPSRC the distribution and annotation of ethical information represented more difficulties, with a list of high-level ethical considerations accompanied by hyperlinks to the RCUK framework and a variety of external sources, many with little or no annotation. One exception is the '*Framework for Ethical and Responsible Innovation'*, which arose from an EPSRC-funded research project and is referenced repeatedly on the website, although its full text is only accessible via a hyperlink to the authors' journal proof. We recommend action to improve consistency amongst RCUK members in their presentation of ethical guidance, including appropriate content tagging, to avoid confusion and facilitate access to relevant advice for researchers using social media in their studies.

The multiplicity of departmental and institutional ethics committees operating within UK universities and research organisations adds further complexity to this landscape. New empirical studies are needed, to shed light on the ways in which such committees are addressing approval requests for studies involving the reuse of data from social media, including which published guidelines they refer to, whether they have their own written policies for this type of research, and whether disciplinary affiliation affects decision making.

Our review of relevant health-related research indexed in one database was intended as an exploratory scoping exercise and should be regarded as indicative rather than exhaustive. We are currently undertaking a comprehensive, rigorous, multi-database, systematic review of data mining research in health, which will inevitably yield further studies. Nonetheless our current results provide valuable insights into the ethical maturity of research involving social media mining and echo the gaps seen in the guidelines we reviewed. We recommend similar analyses of ethical considerations in published articles from other disciplines where social media data are being mined for research, including computer science, the social sciences, economics, business studies, political science and criminology, to name but a few. Given the growing research activities of major social media providers and businesses, research indexed in the scientific literature may represent only the tip of the iceberg, and finding new ways of obtaining access to commercial research would also be worthwhile, although the monetisation of data insights and intellectual property restrictions will inevitably present barriers.

The scope of our analysis did not extend to legal or regulatory aspects of information governance in the context of social media data, which are designed to control or limit certain forms of research. In contrast, ethical guidelines aim to ensure research integrity, discourage irresponsible or socially unacceptable research conduct and support the prioritisation of studies likely to benefit rather than harm society. Likewise, we did not seek to compare methodological innovations such as automated data mining, social network analysis, machine learning or 'black box' algorithms, which also present challenges around consumer choice, control and privacy (Pasquale, 2015). Comparable analyses conducted from each of these perspectives are warranted.

Conclusions and recommendations

Beyond statements about IRB approval, the generally poor integration of ethical concepts and guidelines within the corpus of published articles we have reviewed suggests low levels of awareness amongst researchers using social media mining in their studies, echoing observations from other areas of 'big data' research (e.g. Metcalf et al., 2017). This is consistent with the wide variability we have observed in the research ethics guidance offered by RCUK members in relation to uses of social media platforms and the data derived from them. Our finding that only one RCUK council (ESRC) directly refers to social media research in its ethical guidance is a cause for concern, given the highly interdisciplinary nature of studies in this area, as illustrated by our analysis of relevant health-related publications.

We recommend further cross-council collaboration to develop shared, interdisciplinary guidelines for the ethical use of social media in research, and specifically research involving the harvesting and reuse of social media data.

In the shorter term, effort should be invested to improve consistency in the presentation, accessibility and comprehensiveness of existing ethical guidance available on the various RCUK websites. For example, we observed that some websites are difficult to navigate and contain highly distributed and poorly connected information on ethics, approval processes and regulation. Adequate literature review to ensure the timely inclusion of relevant guidance from other sources is also required; for example, we came across a guide to ethics in social media research which had emerged from a project part-funded by ESRC and EPSRC but was not mentioned on either of their websites (Evans et al., 2015).

Future RCUK ethics guidelines would also benefit from including a broader range of social media uses, clear criteria for judging projects against a variety of ethical considerations, and pragmatic recommendations for researchers planning to undertake studies involving social media.

Until such meta-guidelines are available, we recommend that UK researchers prioritise the existing guidelines produced by the ESRC, BPS, AoIR and NIHR, alongside the ethical taxonomies we have adapted for this study. We also encourage researchers to explore the wider universe of ethical frameworks emerging nationally and internationally in relation to new forms of data, including those from the OECD (2016), the US Council for Big Data Ethics and Society (Metcalf et al., 2017) and the UK Data Service (Bishop, 2017) as well as emerging initiatives such as the UK Society for Data Miners' plans to develop ethical principles (SocDM, 2017) and primary research exploring the boundaries of public acceptability in the reuse of digital personal data (e.g. Aitken et al., 2016; Williams et al., 2017).

We recommend that UK researchers applying for project funding or permission to undertake studies using social media data should explicitly state which ethics guidelines they have consulted, and we call upon IRBs to integrate this requirement into their approvals documentation. We also call upon authors and editors to ensure that publications describing studies involving social media data clearly state the ethical issues that have been considered during the research and specify the guidelines consulted.

Given the substantial investments made in digital research and data science by the UK government and research councils over the last 5 years, coupled with increased policy attention on responsible research and innovation (European Commission, 2013) and the protection of personal data (European Parliament, 2016), ensuring the robust design and implementation of ethical guidelines for social media research is essential.

We hope that the results of this scoping study will inform the future development of such guidelines in the UK and elsewhere, and catalyse a broader interdisciplinary discussion amongst research councils, institutional ethics boards and researchers themselves.

Abbreviations

AoIR: Association of Internet Researchers
AHRC: Arts and Humanities Research Council
BBSRC: Biotechnology and Biological Sciences Research Council
BPS: The British Psychological Society
EPSRC: Engineering and Physical Sciences Research Council
ESRC: Economic and Social Research Council
IRB: Institutional Review Board
MRC: Medical Research Council
NERC: Natural Environment Research Council
NHS: National Health Service (UK)
NIHR: National Institute for Health Research
RCUK: Research Councils United Kingdom
STFC: Science and Technology Facilities Council

Acknowledgements

None.

Competing Interests

This study was conducted as part of JT's self-funded PhD research project, supervised by CP. JT is also an employee of Ernst and Young Ltd and CP is an RCUK grant holder. Neither organization was involved in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Contributorship

Both authors contributed to study conception, planning, analysis and manuscript writing. JT designed and undertook the searches and the email verification exercise, screened the outputs, extracted the data and classified these according the specified taxonomies, with input from and cross-checking by CP.

Ethical approval

This study adheres to the Research Ethics Policy of the University of Edinburgh Medical School and was approved by its Internal Review Board.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: CP is an RCUK grant holder and collaborator on the ESRC Administrative Data Research Centre for Scotland (grant number ES/L007487/1), the MRC Farr Institute for Health Informatics Research UK (grant number MR/K007017/1) and the EPSRC Science and Practice of Social Machines (grant number EP/J017728/1).

Guarantor

Not Applicable.

Notes

Web links cited within the text were last accessed on October 16th, 2017

1. www.facebook.com
2. www.twitter.com
3. www.youtube.com
4. www.linkedin.com
5. www.patientslikeme.com
6. www.mumsnet.com
7. www.baristaexchange.com
8. https://trello.com
9. www.yammer.com
10. www.ushahidi.com
11. www.zooniverse.org

References

AHRC (2016) *Research Funding Guide*. Available at: http://www.ahrc.ac.uk/documents/guides/research-funding-guide/ (accessed 16 October 2017).

Aitken M, de St Jorre J, Pagliari C, et al. (2016) Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics* 17: 73.

Allem JP, Escobedo P, Chu KH, et al. (2016) Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. *Tob Control* 26: 226–229.

Anderson LS, Bell HG, Gilbert M, et al. (2017) Using social listening data to monitor misuse and nonmedical use of bupropion: a content analysis. *JMIR Public Health Surveill* 3: e6.

Anstead N and O'Loughlin B. (2015) Social media analysis and public opinion: The 2010 UK General Election. *J Comput-Mediated Commun* 20: 204–220.

AoIR (2012) *Ethical Decision-Making and Internet Research*. Report from the AoIR Ethics Working Committee (compiled by A Markham and E Buchanan). Available at: http://www.aoir.org/reports/ethics2.pdf (accessed 16 October 2017).

Arthur C (2010) *2010: The first social media election*. Available at: https://www.theguardian.com/media/2010/apr/30/social-media-election-2010 (accessed 16 October 2017).

Balm J (2014) *Open access and social media: helping science move forwards*. Available at: http://www.evidentlycochrane.net/open-access-social-media-can-help-science-move-forwards/ (accessed 16 October 2017).

BBSRC (2017) *BBSRC Research Grants The Guide*. Available at: http://www.bbsrc.ac.uk/documents/grants-guide/ (accessed 16 October 2017).

Bishop L (2017) *Big data and data sharing: Ethical issues*. Available at: https://www.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf (accessed 16 October 2017).

Bjerglund Andersen N and Söderqvist T. (2012) *Social Media and Public Health Research*. Working Paper/Technical Report. University of Copenhagen. Available at: https://bjerglund.files.wordpress.com/2012/11/final-social-media-and-public-health-research1.pdf (accessed 16 October 2017).

Boseley S (2016) *NHS to scrap single database of patients' medical details*. Available at: https://www.theguardian.com/technology/2016/jul/06/nhs-to-scrap-single-database-of-patients-medical-details (accessed 16 October 2017).

BPS (2012) *Conducting Research on the Internet - Guidelines for ethical practice in psychological research online*. Available at: http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf (accessed 16 October 2017).

Cardiff University (2012) *Collaborative Online Social Media Observatory (COSMOS)*. Available at: http://www.cardiff.ac.uk/research/explore/research-units/collaborative-online-social-media-observatory (accessed 16 October 2017).

Cavazos-Rehg P, Krauss M, Grucza R, et al. (2014) Characterizing the followers and tweets of a marijuana-focused Twitter handle. *J Med Internet Res* 16: e157.

Chen AT, Zhu SH and Conway M. (2015) What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. *J Med Internet Res* 17: e220.

Convery I and Cox D. (2012) A review of research ethics in internet-based research. *Practit Res Higher Educ* 6: 1.

Conway M (2014) Ethical issues in using twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. *J Med Internet Res* 16: e290.

Da'ar OB, Yunus F, Md Hossain N, et al. (2016) Impact of Twitter intensity, time, and location on message lapse of bluebird's pursuit of fleas in Madagascar. *J Infect Public Health* 10: 396–402.

Daniulaityte R, Chen L, Lamy FR, et al. (2016) “When ‘bad’ is ‘good’”: identifying personal communication and sentiment in drug-related tweets. *JMIR Public Health Surveill* 2: e162.

David MB (2004) Internet research: privacy, ethics and alienation: an open source approach. *Internet Res* 14: 323–332.

Deiner MS, Lietman TM, McLeod SD, et al. (2016) Surveillance tools emerging from search engines and social media data for determining eye disease patterns. *JAMA Ophthalmol* 134: 1024–1030.

Denecke K, Krieck M, Otrusina L, et al. (2013) How to exploit twitter for public health monitoring? *Methods Inf Med* 52: 326–339.

Doing-Harris KM and Zeng-Treitler Q. (2011) Computer-assisted update of a consumer health vocabulary through mining of social network data. *J Med Internet Res* 13: e37.

EPSRC (2013) *Framework for Responsible Innovation*. Available at: http://www.epsrc.ac.uk/research/framework/

ESRC (2010) *ESRC Framework for Research Ethics*. Available at: http://www.esrc.ac.uk/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2010/ (accessed 16 October 2017).

ESRC (2015) *Framework for research ethics*. Available at: http://www.esrc.ac.uk/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2015/ (accessed 16 October 2017).

European Commission (2013) *Options for Strengthening Responsible Research and Innovation - Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation*. Available at: http://ec.europa.eu/research/science-society/document_library/pdf_06/options-for-strengthening_en.pdf (accessed 16 October 2017).

European Parliament (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Evans H, Ginnis S and Bartlett J (2015) *#SocialEthics a guide to embedding ethics in social media research*. Available at: https://www.ipsos.com/sites/default/files/migrations/en-uk/files/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf (accessed 16 October 2017).

Eysenbach G and Till JE (2001) Ethical issues in qualitative research on internet communities. *BMJ* 323: 1103.

Fromm J (2016) *New Study Finds Social Media Shapes Millennial Political Involvement and Engagement*. Forbes. June 22nd. https://www.forbes.com/sites/jefffromm/2016/06/22/new-study-finds-social-media-shapes-millennial-political-involvement-and-engagement/#35cf98e22618 (accessed 16 October 2017).

Giles EL, Holmes M, McColl E, et al. (2015) Acceptability of financial incentives for breastfeeding: thematic analysis of readers’ comments to UK online news reports. *BMC Pregnancy Childbirth* 15: 116.

Good BM, Clarke EL, Loguercio S, et al. (2012) Building a biomedical semantic network in Wikipedia with Semantic Wiki Links. *Database (Oxford)* 2012: bar060.

Grant A (2016) “#discrimination”: the online response to a case of a breastfeeding mother being ejected from a UK retail premises. *J Hum Lact* 32: 141–151.

Gregg R, Patel A, Patel S, et al. (2017) Public reaction to the UK government strategy on childhood obesity in England: A qualitative and quantitative summary of online reaction to media reports. *Health Policy* 121: 450–457.

Gu H, Chen B, Zhu H, et al. (2014) Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. *J Med Internet Res* 16: e20.

Guan L, Hao B, Cheng Q, et al. (2015) Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. *JMIR Ment Health* 2: e17.

Heaivilin N, Gerbert B, Page JE, et al. (2011) Public health surveillance of dental pain via Twitter. *J Dent Res* 90: 1047–1051.

Hearne E and Van Hout MC (2016) “Trip-Sitting” in the black hole: a netnographic study of dissociation and indigenous harm reduction. *J Psychoactive Drugs* 48: 233–242.

Hilyard K, Broniatowski D and Dredze M. (2015) *How Far Can Twitter Reach in Good Survey Research?* Available at: http://www.socialsciencespace.com/2015/04/how-far-can-twitter-reach-in-good-survey-research/ http://www.socialsciencespace.com/2015/04/how-far-can-twitter-reach-in-good-survey-research/ (accessed 16 October 2017).

Hokby S, Hadlaczky G, Westerlund J, et al. (2016) Are mental health effects of internet use attributable to the web-based content or perceived consequences of usage? A longitudinal study of European adolescents. *JMIR Ment Health* 3: e31.

Kemp S (2017) Digital in 2017: Global Overview. We Are Social and Hootsuite. Online. Available at https://wearesocial.com/special-reports/digital-in-2017-global-overview (accessed 16 October 2017).

Huang J, Kornfield R, Szczypka G, et al. (2014) A cross-sectional examination of marketing of electronic cigarettes on Twitter. *Tob Control* 23(Suppl. 3): iii26–30.

Huh J, Yetisgen-Yildiz M and Pratt W (2013) Text classification for assisting moderators in online health communities. *J Biomed Inform* 46: 998–1005.

INVOLVE. (2014) *Guidance on the use of social media to actively involve people in research*. Available at: http://www.invo.org.uk/wp-content/uploads/2014/11/9982-Social-Media-Guide-WEB.pdf (accessed 16 October 2017).

Jha A, Lin L and Savoia E (2016) The use of social media by state health departments in the US: analyzing health communication through Facebook. *J Community Health* 41: 174–179.

Jouhki J, Lauk E, Penttinen M, Sormanen N and Uskali T. (2016). Facebook’s Emotional Contagion Experiment as a Challenge to Research Ethics. *Media and Communication*, 4 (4), 75-85. doi:10.17645/mac.v4i4.579

Kaplan A and Haenlein M (2010) Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 53: 59–68.

Karimi S, Metke-Jimenez A, Kemp M, et al. (2015) Cadec: A corpus of adverse drug event annotations. *J Biomed Inform* 55: 73–81.

Kendal S, Kirk S, Elvey R, et al. (2017) How a moderated online discussion forum facilitates support for young people with eating disorders. *Health Expect* 20: 98–111.

Kiss J (2014) *An online Magna Carta: Berners-Lee calls for bill of rights for web*. Available at: http://www.theguardian.com/technology/2014/mar/12/online-magna-carta-berners-lee-web (accessed 16 October 2017).

Knapton S (2014) *Health records of every NHS patient to be shared in vast database*. Available at: http://www.telegraph.co.uk/news/10565160/Health-records-of-every-NHS-patient-to-be-shared-in-vast-database.html (accessed 16 October 2017).

Koene A and Adolphs S (2015) *Ethics considerations for Corpus Linguistic studies using internet resources*. Available at: http://casma.wp.horizon.ac.uk/wp-content/uploads/2015/04/CL2015-CorpusLinguisticsEthics_KoeneAdolphs.pdf (accessed 16 October 2017).

Leggatt-Cook C and Chamberlain K (2012) Blogging for weight loss: personal accountability, writing selves, and the weight-loss blogosphere. *Sociol Health Illn* 34: 963–977.

Liu S, Zhu M, Yu DJ, et al. (2017) Using real-time social media technologies to monitor levels of perceived stress and emotional state in college students: a web-based questionnaire study. *JMIR Ment Health* 4: e2.

Mackey TK, Eysenbach G, Liang BA, et al. (2014) A call for a moratorium on the .health generic top-level domain: preventing the commercialization and exclusive control of online health information. *Globaliz Health* 10: 62.

Mandl KD, McNabb M, Marks N, et al. (2014) Participatory surveillance of diabetes device safety: a social media-based complement to traditional FDA reporting. *J Am Med Inform Assoc* 21: 687–691.

Marshall SA, Yang CC, Ping Q, et al. (2016) Symptom clusters in women with breast cancer: an analysis of data from social media and a research study. *Qual Life Res* 25: 547–557.

Matsuda S, Aoki K, Tomizawa S, et al. (2017) Analysis of patient narratives in disease blogs on the internet: an exploratory study of social pharmacovigilance. *JMIR Public Health Surveill* 3: e10.

McKee HA (2011) Policy matters now and in the future: net neutrality, corporate data mining, and government surveillance. *Computers and Composition* 28: 276–291.

McKee R (2013) Ethical issues in using social media for health and health care research. *Health Policy* 110: 298–301.

Meaney S, Cussen L, Greene RA, et al. (2016) Reaction on Twitter to a cluster of perinatal deaths: a mixed method study. *JMIR Public Health Surveill* 2: e36.

Metcalf J, Keller EF and Boyd D. (2017) *Perspectives on Big Data, Ethics, and Society*. Available at: http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/ (accessed 16 October 2017).

Moorhead SA, Hazlett DE, Harrison L, et al. (2013) A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res* 15: e85.

MRC (2000) *Personal Information in Medical Research*. Available at: http://www.mrc.ac.uk/documents/pdf/personal-information-in-medical-research/ (accessed 16 October 2017).

MRC (2012a) *Good research practice*. Available at: http://www.mrc.ac.uk/news-events/publications/good-research-practice-principles-and-guidelines/ (accessed 16 October 2017).

MRC (2012b) *MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies*. Available at: https://www.mrc.ac.uk/publications/browse/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/ (accessed 16 October 2017).

Mudry TE and Strong T (2013) Doing recovery online. *Qual Health Res* 23: 313–325.

Munson SA, Cavusoglu H, Frisch L, et al. (2013) Sociotechnical challenges and progress in using social media for health. *J Med Internet Res* 15: e226.

Murphy J, Link MW, Hunter Childs J, et al. (2014) *Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research*. Available at: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf (accessed 16 October 2017).

National Data Guardian (2017) *National Data Guardian (NDG) statement on government response to the NDG Review*. Available at: https://www.gov.uk/government/news/national-data-guardian-ndg-statement-on-government-response-to-the-ndg-review (accessed 16 October 2017).

NERC (2015) *Ethics Policy*. Available at: http://www.nerc.ac.uk/about/policy/policies/nerc-ethics-policy.pdf (accessed 16 October 2017).

Nunan D and Yenicioglu B (2013) Informed, uninformed and participative consent in social media research. *Int J Market Res* 55: 791.

O'Sullivan B (2017) *Towards a Magna Carta for Data*. Dublin: Royal Irish Academy.

Odlum M and Yoon S (2015) What can we learn about the Ebola outbreak from tweets? *Am J Infect Control* 43: 563–571.

OECD (2016) *Research Ethics and New Forms of Data for Social and Economic Research*: OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/research-ethics-and-new-forms-of-data-for-social-and-economic-research_5jln7vnpxs32-en (accessed 16 October 2017).

Ofoghi B, Mann M and Verspoor K. (2016) Towards early discovery of salient health threats: a social media emotion classification technique. *Pac Symp Biocomput* 21: 504–515.

Orton-Johnson K (2010) Ethics in online research; evaluating the ESRC framework for research ethics categorisation of risk. *Sociol Res Online* 15: 13.

Pagliari C and Vijaykumar S (2016) Digital Participatory Surveillancce and the Zika Crisis: Opportunities and Caveats. *PLoS Negl Trop Dis* 10(6) e0004795

Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.

Paul MJ and Dredze M (2011) *You Are What You Tweet: Analyzing Twitter for Public Health*. Available at: http://www.cs.jhu.edu/~mpaul/files/2011.icwsm.twitter_health.pdf (accessed 16 October 2017).

Paul MJ and Dredze M. (2014) Discovering health topics in social media using topic models. *PLoS One* 9: e103408.

Pedersen ER and Kurz J (2016) Using Facebook for health-related research study recruitment and program delivery. *Curr Opin Psychol* 9: 38–43.

Pells R (2017) *Facebook research targeted insecure youth, leaked documents show*. Available at: http://www.independent.co.uk/news/media/facebook-leaked-documents-research-targeted-insecure-youth-teenagers-vulnerable-moods-advertising-a7711551.html (accessed 16 October 2017).

Powell GE, Seifert HA, Reblin T, et al. (2016) Social media listening for routine post-marketing safety surveillance. *Drug Saf* 39: 443–454.

RCUK (2013) *RCUK Policy and Guidelines on Governance of Good Research Conduct*. Available at: http://www.rcuk.ac.uk/documents/reviews/grc/rcukpolicyguidelinesgovernancegoodresearchconduct-pdf/ (accessed 16 October 2017).

Richardson A and Vallone DM. (2014) YouTube: a promotional vehicle for little cigars and cigarillos? *Tob Control* 23: 21–26.

Risson V, Saini D, Bonzani I, et al. (2016) Patterns of treatment switching in multiple sclerosis therapies in US patients active on social media: application of social media content analysis to health outcomes research. *J Med Internet Res* 18: e62.

Rodrigues RG, das Dores RM, Camilo-Junior CG, et al. (2016) SentiHealth-Cancer: a sentiment analysis tool to help detecting mood of patients in online social networks. *Int J Med Inform* 85: 80–95.

Russell DJ, Sprung J, McCauley D, et al. (2016) Knowledge exchange and discovery in the age of social media: the journey from inception to establishment of a parent-led web-based research advisory community for childhood disability. *J Med Internet Res* 18: e293.

Sarker A, O'Connor K, Ginn R, et al. (2016) Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug Saf* 39: 231–240.

Schomberg JP, Haimson OL, Hayes GR, et al. (2016) Supplementing public health inspection via social media. *PLoS One* 11: e0152117.

Seidenberg AB, Pagoto SL, Vickey TA, et al. (2016) Tanning bed burns reported on Twitter: over 15,000 in 2013. *Transl Behav Med* 6: 271–276.

Sinclair W, McLoughlin M and Warne T. (2015) To Twitter to Woo: harnessing the power of social media (SoMe) in nurse education to enhance the student's experience. *Nurse Educ Pract* 15: 507–511.

Smith C (2014) *Social Networks Are Only Just Getting Started In Mining User Data*. Available at: http://www.businessinsider.com/social-medias-big-data-future-2014-2?IR=T http://www.rcuk.ac.uk/documents/reviews/grc/rcukpolicyguidelinesgovernancegoodresearch-conduct-pdf/ (accessed 16 October 2017).

SocDM (2017) *Society of Data Miners: Towards a Professional Code of Ethics for Data Mining*. Available at: https://www.turing.ac.uk/events/socdm-towards-professional-code-ethics-data-mining-part-2/ (accessed 16 October 2017).

Song J, Song TM, Seo DC, et al. (2017) Social big data analysis of information spread and perceived infection risk during the 2015 Middle East respiratory syndrome outbreak in South Korea. *Cyberpsychol Behav Soc Netw* 20: 22–29.

Song J, Song TM, Seo DC, et al. (2016) Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents. *J Adolesc Health* 59: 668–673.

STFC (2013) *Public Engagement with Science and Technology Strategic Plan 2013-2016*. Available at: http://webarchive.nationalarchives.gov.uk/20150106104755/https://www.stfc.ac.uk/1780.aspx (accessed 16 October 2017).

Strekalova YA (2017) Health risk information engagement and amplification on social media. *Health Educ Behav* 44: 332–339.

Sudau F, Friede T, Grabowski J, et al. (2014) Sources of information and behavioral patterns in online health forums: observational study. *J Med Internet Res* 16: e10.

Trisha G (2013) Citizen science: amateur experts. *Nature* 496: 259.

Tsuya A, Sugawara Y, Tanaka A, et al. (2014) Do cancer patients tweet? Examining the twitter use of cancer patients in Japan. *J Med Internet Res* 16: e137.

Tursunbayeva A, Franco M and Pagliari C (2017). Use of social media for e-Government in the public health sector: A systematic review of published studies. Government Information Quarterly. DOI: 10.1016/j.giq.2017.04.001

Van Hout MC and Hearne E (2015) "Vintage meds": a netnographic study of user decision-making, home preparation, and consumptive patterns of laudanum. *Subst Use Misuse* 50: 598–608.

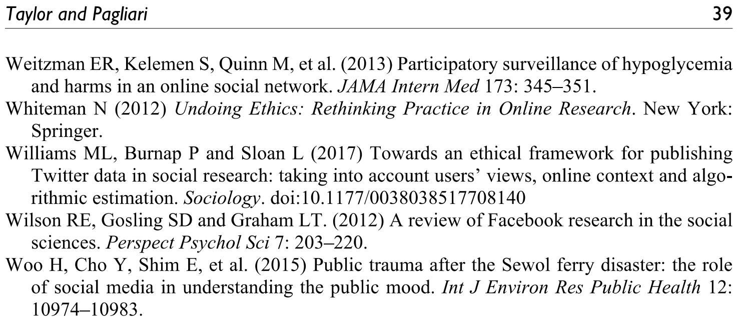
van Mierlo T, Li X, Hyatt D, et al. (2017) Demographic and indication-specific characteristics have limited association with social network engagement: evidence from 24,954 members of four health care support groups. *J Med Internet Res* 19: e40.

Vayena E, Mastroianni A and Kahn J (2012) Ethical issues in health research with novel online sources. *Am J Public Health* 102: 2225–2230.

Wakefield J (2017) *Google DeepMind's NHS deal under scrutiny*. Available at: http://www.bbc.com/news/technology-39301901 (accessed 16 October 2017).

Weitzman ER, Adida B, Kelemen S, et al. (2011a) Sharing data for public health research by members of an international online diabetes social network. *PLoS One* 6: e19256.

Weitzman ER, Kelemen S and Mandl KD (2011b) Surveillance of an online social network to assess population-level diabetes health status and healthcare quality. *Online J Public Health Inform* 3(3): ojphi.v3i3.3797.



Weitzman ER, Kelemen S, Quinn M, et al. (2013) Participatory surveillance of hypoglycemia and harms in an online social network. *JAMA Intern Med* 173: 345–351.
Whiteman N (2012) *Undoing Ethics: Rethinking Practice in Online Research*. New York: Springer.
Williams ML, Burnap P and Sloan L (2017) Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. *Sociology*. doi:10.1177/0038038517708140
Wilson RE, Gosling SD and Graham LT. (2012) A review of Facebook research in the social sciences. *Perspect Psychol Sci* 7: 203–220.
Woo H, Cho Y, Shim E, et al. (2015) Public trauma after the Sewol ferry disaster: the role of social media in understanding the public mood. *Int J Environ Res Public Health* 12: 10974–10983.

A.2.2.1 Additional information and material

Media coverage:

Daniel Sanderson, 'Studies using social media 'ignore ethics'', The Times, page 2, 28 October 2017.
https://www.thetimes.co.uk/article/studies-using-social-media-ignore-ethics-bqqpcpv3g

University of Edinburgh. 'Social media data use needs tighter research controls, experts say', ScienceDaily. ScienceDaily, 30 October 2017.
www.sciencedaily.com/releases/2017/10/171030134628.htm

University of Edinburgh. 'Social media data use needs tighter research controls, experts say', EurekAlert!, EurekAlert! Science News, 30 October 2017.
https://www.eurekalert.org/pub_releases/2017-10/uoe-smd103017.php

Cara Livernois, 'Social media research data often lacks ethical guidelines, need recommendations', Clinical Innovation + Technology, 30 October 2017.
https://www.aiin.healthcare/topics/research/social-media-research-data-often-lacks-ethical-guidelines-need-recommendations

PhysOrg. 'Social media data use needs tighter research controls, experts say', 30 October 2017.
https://phys.org/news/2017-10-social-media-tighter-experts.html

A.2.3 npj Digital Medicine (published June 2019)

Altmetric attention score: 55 (as of 2 May 2020)
Number of citations: 2 (as of 2 May 2020)

npj | Digital Medicine
www.nature.com/npjdigitalmed

ARTICLE **OPEN**

The social dynamics of lung cancer talk on Twitter, Facebook and Macmillan.org.uk

Joanna Taylor[1] and Claudia Pagliari[1]

People with lung cancer and others affected by the condition are using social media to share information and support, but little is known about how these behaviours vary between different platforms. To investigate this, we extracted posts from Twitter (using relevant hashtags), the Lung Cancer Support Group on Facebook and the Macmillan.org.uk lung cancer discussion forum for a single month. Interaction Process Analysis revealed that all three platforms were used more for giving than seeking information, opinion or suggestions. However, interaction types (including sentiment) varied between platforms, reflecting their digital architectures, user-base and inclusion of a moderator. For example, a higher percentage of information-seeking and sentiment marked the Macmillan.org.uk, compared with Twitter and the Facebook Group. Further analysis of the messages using a four-dimensional typology of social support revealed that emotional and informational support types were most prevalent on the Macmillan.org.uk forum, closely followed by the Facebook Group. Contrary to expectations, Twitter posts showed the most companionship support, reflecting the use of hashtags as user-generated signals of community belonging and interests. Qualitative analysis revealed an unanticipated sub-category of spiritual support, which featured uniquely in the Lung Cancer Support Group on Facebook. There was little evidence of trolling or stigma, although some users remarked that lung cancer was unfairly resourced compared with other cancers. These findings provide new insights about how people affected by lung cancer use social media and begin to elucidate the value of different platforms as channels for patient engagement and support, or as potential research data sources.

npj Digital Medicine (2019)2:51 ; https://doi.org/10.1038/s41746-019-0124-y

INTRODUCTION

Lung cancer is the most common cancer globally, with ~2.09 million cases every year.[1] There are two main types of primary lung cancer: small cell lung cancer (SCLC), affecting 15% of those diagnosed in the United States, and non-SCLC (NSCLC) affecting 85%.[2] The severity of this condition varies between Stage 1 and Stage 4, depending on the size of the tumour and whether it has spread. These stages can influence survival rates with fewer than 6% of patients living more than 5 years after diagnosis with Stage 3. Treatment options include but are not limited to surgery, chemotherapy, radiotherapy and laser therapy, and are dependent on the individual. Causes of lung cancer include smoking, passive smoking, exposure to radon and asbestos, air pollution, low immunity and family history, to name but a few.[3] Although smoking is a known risk factor, 10–15% of people who develop lung cancer are never-smokers and its cause cannot be definitively associated with established environmental risk factors.[4] Much research has been conducted into the stigma associated with lung cancer[5–7] and has shown that it is considered more highly stigmatized than other cancers[8,9] due to self-blame and its causal attribution to smoking.[10,11] As such, those diagnosed with lung cancer are encouraged to seek support through support groups and online communities, in order to reduce the likelihood of depression.[12]

Social media are online, often mobile, platforms that support the creation and exchange of user-generated content.[13] They are estimated to have 2.46 billion users worldwide.[14] Hamm et al.'s[15] scoping review of studies involving social media use by patients and caregivers reported that discussion forums, online support groups, social networking sites and micro-blogs dominate the research literature, with 11.3% of the identified studies focusing on cancer. Patel et al.'s[16] systematic review of research on the use of social media in chronic disease, as defined by the Centers for Disease Control, went further by evaluating the clinical outcomes of such technologies. This revealed that Facebook, blogs and Twitter were the most popular social media examined, and that cancer was the most common chronic condition investigated. Although relevant research is fragmented and currently lacking in substantive empirical evidence, existing studies suggest that social media can be used to provide social, emotional or experiential support in chronic disease management and are likely to improve patient care.[16,17] Psychological support was revealed to be present in the majority of tweets by cancer patients,[18] whereas a narrative synthesis of cancer patient blogs indicated that users share their diagnosis and treatment journeys online as a means of describing their experiences of health services, informing their health behaviour and in maintaining relationships with others.[19] Lung cancer is the second most prevalent cancer discussed on Twitter, after breast cancer,[20] and research has revealed that the majority of relevant tweets focus on treatment and the use of pharmaceutical and research interventions, followed by awareness-raising and prevention/risks.[21]

Health researchers are increasingly using data sourced from social media to understand how members of patient communities

---

[1] eHealth Research Group, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, UK
Correspondence: Joanna Taylor (joanna.taylor@ed.ac.uk)

Received: 14 October 2018 Accepted: 15 May 2019
Published online: 10 June 2019

Table 1. Summary of social media posts from October 2017

Source	Number of unique authors	Total sample size	Number of original posts	Number of replies	Number of secondary replies	Number of posts included in IPA categorization stage (%)
Twitter #LCSM and #LungCancer	1056	3000	3000	0	0	2897 (97%)
Facebook Lung Cancer Support Group	844	2644	51	1659	934	2597 (98%)
Macmillan.org.uk lung cancer discussion forum	96	266	51	215	0	266 (100%)

interact with each other regarding specific conditions.[22,23] Single platforms, such as Twitter and Facebook or condition-specific online communities, have dominated previous research. However, a study comparing the use of different social media platforms by patients with Type 1 diabetes revealed variations in the purposes for which these were used, with Twitter mainly used for information and opinion sharing, with little support or empathy, and discussion forums and social networking sites used more often for social interaction and peer support.[24]

We expand on the latter study here, by exploring the types of interaction and support demonstrated on different social media platforms by people affected by lung cancer, with specific reference to lung cancer hashtags on Twitter,[25] the Lung Cancer Support Group on Facebook[26] and the lung cancer discussion forum on Macmillan.org.uk.[27] In doing so, we consider the following research questions: (1) Do people affected by lung cancer use different social media in different ways? and (2) which social media are most successful at encouraging social interaction and support for people affected by lung cancer?

RESULTS

Frequency of usage

Table 1 shows the number of lung cancer-related posts extracted from each of the three social media platforms, the total number of people contributing to each platform, the number of replies and the number of relevant English language posts included in the categorization stage. The Twitter hashtags #LCSM and #LungCancer were found to have the highest absolute number of posts (3000 posts over the 1-month period), followed by the Lung Cancer Support Group on Facebook (2644 posts) and the lung cancer discussion forum on Macmillan.org.uk (266 posts).

Contributors

The Twitter hashtags had the highest number of unique authors (1056), compared with the Lung Cancer Support Group on Facebook (844) and the lung cancer discussion forum on Macmillan.org.uk (96). At 2593, the Lung Cancer Support Group on Facebook had the highest number of replies compared with the Macmillan lung cancer discussion forum (215 replies) and the Twitter hashtags (0 replies).

Types of interaction

Table 2 and Fig. 1 show the frequency of posts fitting each of the 12 categories developed by Bales[28] (Fig. 2) for classifying the type (not content) of interactions taking place in groups, known as Interaction Process Analysis (IPA). Although all three platforms were mainly used to post suggestions, opinions and information, information sharing was the most common use, representing 64% of posts bearing the Twitter lung cancer hashtags, 58% of posts to Macmillan's lung cancer discussion forum and 43% of posts to the Lung Cancer Support Group on Facebook. Noticeably, fewer posts sought suggestions, opinions and information from other members, although this varied across platforms (5% of Twitter lung cancer hashtag posts, 7% of posts in the Lung Cancer Support Group on Facebook and 28% of the posts on Macmillan's lung cancer discussion forum). There were also more posts classified as friendly, unfriendly, tension release and showing tension in the sample from Macmillan's lung cancer discussion forum (56.8% friendly, 6%, shows tension, 5.6% tension release, 1.1% unfriendly) compared with the Lung Cancer Support Group on Facebook (37.5% friendly, 1.7% shows tension, 0.7% unfriendly, 0.4% tension release) and the Twitter lung cancer hashtags (11.3% friendly, 0.6% unfriendly, 0.4% shows tension release, 0.2% tension release), suggesting a greater degree of sentiment expressed in the Macmillan discussion forum. To provide transparency and increase the reproducibility of our analysis, examples of posts from each category and social media platform are provided in Table 3.

Types of social support

After excluding promotional or news posts (45.1% of the Twitter sample only), data from the three platforms was analysed against the four-dimensional typology of social support.[29–31] The four social support types are Emotional (offering empathy, concern, affection, love, trust, acceptance, intimacy, encouragement or caring), Instrumental (provision of financial assistance, material goods, services or tangible aid), Informational (provision of advice, guidance, suggestions or useful information to someone), and Companionship (design to give a sense of belonging). The results are summarized in Table 4. These highlight differences between the three social media sources, with informational support being more evident on the discussion forum on Macmillan.org.uk (65.4%) and Lung Cancer Support Group on Facebook (54.7%) compared with the Twitter hashtags (29.1%). Emotional support is also most evident in the lung cancer discussion forum on Macmillan.org.uk (66.9%) compared with Lung Cancer Support Group on Facebook (51.3%) and the Twitter hashtags (5.8%).

Message content and sentiment

In lieu of qualitative analysis, the frequency and co-occurrence of keywords associated with posts in each social support category, from the three social media platforms, were mapped into the semantic 'word clouds' shown in Fig. 3. No word cloud was produced for 'Instrumental' support, due to the small number of posts in this category. The themes characterizing posts falling into each of the remaining three social support categories are shown in the right-hand column. Emotional support is represented by qualitative themes such as spirituality, grief, family and positive sentiment, whereas informational support is identified by terms related of the diagnosis and treatment of the condition. Keywords relevant to community and advocacy are evident in the category of companionship support.

Table 2. Application of Bales IPA to the lung cancer posts

Source	Total number of posts analysed	Seems friendly	Shows tension release	Agrees	Asks for suggestion	Asks for opinion	Asks for information	Gives information	Gives opinion	Gives suggestion	Disagrees	Shows tension	Seems unfriendly
Twitter #LCSM and #LungCancer	2897	326 (11.3%)	6 (0.2%)	39 (1.3%)	2 (0.1%)	55 (1.9%)	101 (3.5%)	1849 (63.9%)	567 (19.6%)	453 (15.6%)	1 (0.03%)	11 (0.4%)	18 (0.6%)
Facebook Lung Cancer Support Group	2597	973 (37.5%)	10 (0.4%)	105 (4.0%)	1 (0.04%)	45 (1.7%)	149 (5.7%)	1117 (43.0%)	710 (27.3%)	231 (8.9%)	0 (0%)	43 (1.7%)	17 (0.7%)
Macmillan.org.uk lung cancer discussion forum	266	151 (56.8%)	15 (5.6%)	4 (1.4%)	6 (2.3%)	24 (9.0%)	44 (16.5%)	154 (57.9%)	90 (33.8%)	51 (19.2%)	0 (0%)	16 (6.0%)	3 (1.1%)

The categories were not considered mutually exclusive

DISCUSSION

This descriptive analysis, encompassing a total of 5910 relevant messages posted on Twitter using the hashtags #LCSM and #LungCancer, the Lung Cancer Support Group on Facebook and the Macmillian.org.uk lung cancer discussion forum, identified variations in the use of different social media by people affected by lung cancer. These variations include the nature of interactions within these online communities and the type of social support represented. Across the three platforms, the absolute number of postings in the 1-month observation period was greatest for the Twitter hashtags, followed by the Lung Cancer Support Group on Facebook and then the lung cancer discussion forum on Macmillan.org.uk. However, these raw numbers say little about their value for users.

As noted earlier, the Bales' IPA highlighted similarities and differences in the type of interactions found on each of the three social media platforms. All are predominantly used for information sharing, with the lung cancer discussion forum on Macmillan.org.uk also showing the most posts expressing sentiment (e.g., friendliness, tension).

There are several possible explanations for these differences. One relates to the digital architectures of the platforms—defined as 'the technical protocols that enable, constrain and shape user behaviour in a virtual space'.[32] Twitter is designed for individual micro-blogging (broadcasting), whereas Facebook groups and discussion forums are designed for conversation and sharing. Although Twitter posts may trigger chains of responses, this is coincidental, whereas Facebook groups are explicitly designed for this and include invited members, often mirroring offline social networks.[33] In our study, high levels of two-way communication were seen in the Lung Cancer Support Group on Facebook (99.8%) and lung cancer discussion forum on Macmillan.org.uk (81%), whereas the sample of Twitter posts contained only original tweets and retweets.

An observation that warrants further investigation, relates to the types of account holders posting to the different social media platforms. In previous analyses of health-related Twitter narratives, 25% of verified accounts belonged to journalists,[33] 40% to companies and brand accounts, and 15–20% were bots.[34] In our study, organizations (e.g., pharmaceutical companies, charities), patient advocacy groups, research institutions and news outlets appear to use Twitter to disseminate health information to the general public. In contrast, the Lung Cancer Support Group on Facebook and the discussion forum on Macmillan.org.uk were chiefly designed for people affected by lung cancer and due to the presence of community moderators, function as online support groups and enable a greater degree of interaction.

Community moderators or administrators play an influential role in online communities,[35,36] as was the case in the closed (members only) Lung Cancer Support Group on Facebook and the lung cancer discussion forum on Macmillan.org.uk where their presence created greater understanding of the community rules and expectations. In the case of the Lung Cancer Support Group on Facebook, group members were asked to agree to a code of conduct, requiring them to respect others, avoid foul language, focus on lung cancer, not seek medical advice, refrain from advertising or fund-raising and never block an admin, or face removal. As well as providing a 'safe' place for people to share information and experiences, moderators of the Macmillan.org.uk lung cancer discussion forum responded to posts that had not yet received a comment from other community members. Examples of this include 'sorry to hear your news, and I'm also sorry you've not had a reply yet'. In responding to this post, the moderator was able to increase the visibility of the post in an attempt to garner a response. In contrast, moderation of content posted on Twitter is absent, except during the 1 hour pre-scheduled #LCSM group discussion that takes place every 2 weeks.[37] The moderator in

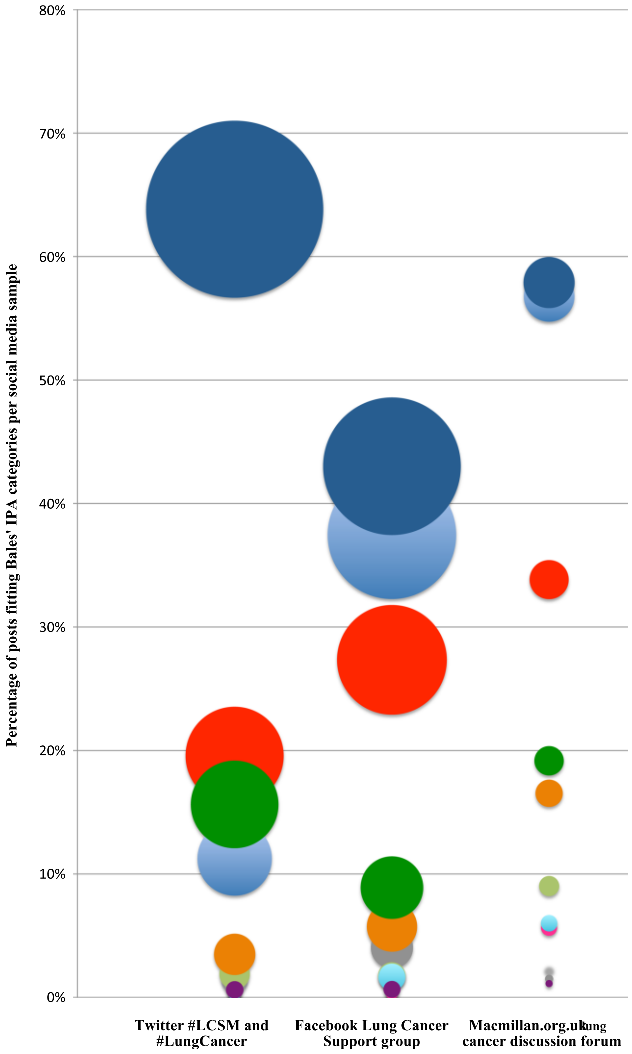


Fig. 1 Percentage of posts in the corpus of data from the Twitter lung cancer hashtags, the Lung Cancer Support Group on Facebook and the lung cancer discussion forum on Macmillian.org.uk, and their fit to Bales' IPA categories. The size of each circle represents the percentage of posts associated with each of Bales' IPA categories. Each colour represents a different category: light blue for 'seems friendly', fuchsia for 'shows tension release', grey for 'agrees', black for 'asks for suggestion', light green for 'asks for opinion', orange for 'asks for information', dark blue for 'gives information', red for 'gives opinion, dark green for 'gives suggestion', lilac for 'disagrees', turquoise for 'shows tension' and purple for 'seems unfriendly'

these cases is there to greet participants and facilitate a discussion based on predefined questions (e.g., 'I will announce four topic questions (T1 T2 etc). Pls label your answers with T1, T2, etc to make transcript easier to follow #lcsm').

Findings also revealed that posts by any user that are similar to an online survey would generate the most responses on Facebook, as was also the case in a previous diabetes study.[24] Examples of this include: 'how old was everyone when they were

Positive reactions	1	*Seems friendly,* shows solidarity, raises other's status, gives help, reward
	2	*Shows tension release,* jokes, laughs, shows satisfaction
	3	*Agrees,* shows passive acceptance, understands, concurs, complies
Attempted answers	4	*Gives suggestion,* direction, implying autonomy for others
	5	*Gives opinion,* evaluation, analysis, expresses feeling, wish
	6	*Gives information,* orientation, repeats, clarifies, confirms
Questions	7	*Asks for information,* orientation, repetition, confirmation
	8	*Asks for opinion,* evaluation, analysis expression of feeling
	9	*Asks for suggestion,* direction, possible ways of action
Negative reactions	10	*Disagrees,* shows passive rejection, formality, withholds help
	11	*Shows tension,* asks for help, withdraws out of field
	12	*Seems unfriendly,* shows antagonism, deflates other's status, defends or asserts self

Fig. 2 Description of each of Bales' IPA categories

first diagnosed?' (609 responses), 'this may be a really dumb question: can someone survive lung cancer?' (228 responses) and 'anyone got a husband or wife that won't give up smoking?' (73 responses).

In summary, although the Lung Cancer Support Group on Facebook and lung cancer discussion forum on Macmillan.org.uk were moderated, this is limited in Twitter to pre-scheduled group discussions. We find that Twitter was the most actively used social media in terms of volume of posts, whereas Facebook achieved the highest percentage response rate and interaction. The differences in digital architecture, in turn contribute to the variations in social interaction and support for people affected by lung cancer.

We observed several interesting differences in the social interaction represented in the posts on these three social media platforms. As previously noted, the lung cancer hashtags on Twitter are mainly used to disseminate information (64%) and opinion (20%) in line with research describing Twitter as a mass communication and broadcasting tool[38] with the majority of active narratives involving two or fewer users.[39] Findings from this study are similar in nature to that of the Type 1 diabetes comparative study from which we drew inspiration.[24]

The positive correlation between social support and health is widely accepted among the public health and psychology research communities.[40] In addition to types of functional interaction shown in the posts, our analysis revealed differences in the types of social support that are being sought and offered by users of different social media. All content posted in the Macmillan.org.uk lung cancer discussion forum and in the Lung Cancer Support Group on Facebook was associated with at least one of the functions of social support whilst 43.6% of tweets were not and in most cases can be considered either news reports (e.g., 'FDA Approves Higher-Dose Tablet of Brigatinib for NSCLC'), promotional messages (e.g., 'Get your #lungcancer swag for #LCAM17 this Nov. T-shirts, tank tops, sunglasses, bracelets and more!') or advocacy-related (e.g., '#433aday Lung cancer kills 433 Americans a day. We need better funding for research now') in content.

Informational support, as already noted, is the provision of information aimed at supporting a member or members of the social network, often in response to a statement of distress or a request for help. Although this overlaps with Bales' IPA categories of 'information giving' they are not synonymous. The lung cancer discussion forum on Macmillan.org.uk produced the highest percentage of informational support posts (65.7%), followed by the Lung Cancer Support Group on Facebook (54%), the #LungCancer on Twitter (40.7%) and the #LCSM on Twitter (15.4%). Posts such as 'Looking for advice. Mum is in later stages of stage 4 lung cancer….In the last 4 days mum has stopped eating and can only get a small amount of fluids in her…Not sure what to do or what to expect now' and 'can radio be used after Keytruda? Can it be keytruda + radio?' illustrate the type of informational support sought. Posts classified as informational support are represented in words related to the diagnosis, treatment and progression of the condition over time, a finding which aligns somewhat to that from Tsuya et al.[19] study into whether cancer patients tweet.

The lung cancer discussion forum on Macmillan.org.uk produced the highest percentage of emotional support posts (67.2%), followed by the Lung Cancer Support Group on Facebook (50.6%), the lung cancer hashtags on Twitter (5.8%), contradicting findings from previous research that revealed that the majority of tweets posted by cancer patients focused on psychological support.[18] Emotional support is represented in posts such as 'I am so sorry to hear this…you are in my thoughts' and 'It really is so hard…I'm full of hurt and anger. Just seems so unfair'. Words associated with emotional support include spiritual and religious terms (e.g., faith, god, hope, prayers), grief (e.g., sorry, loss), family and positive sentiment (e.g., thanks, hugs).

Given the severity of lung cancer, its treatment and the life-threatening nature of the condition, it is not surprising that spiritual and existential beliefs are represented in the social media data[41]; Lung Cancer Support Group on Facebook (14.5%), lung cancer discussion forum on Macmillan.org.uk (3.4%) and lung cancer hashtags on Twitter (0.7%). Identified as one of the modifiable dimensions of the patient experience, much research has been conducted into the role spirituality and faith plays in the illness trajectories of lung cancer patients.[42–44] Its manifestation in social media, however, is a relatively new field[45–47] with little empirical evidence of how social media platforms differ in this context. Findings from our study suggest that spirituality is significantly more prevalent on Facebook rather than Twitter and further research is warranted to test this hypothesis.

Surprisingly, the Twitter lung cancer hashtags indicated the highest percentage of companionship support (23.7%) compared with the lung cancer discussion forum on Macmillan.org.uk (12.5%) and Lung Cancer Support Group on Facebook (2.7%). Words associated with this form of social support include those related to joining online conversations and physical events, manifesting themself in posts such as 'Sending so much love to

Table 3. Examples of the different categories of post

Source	Seems friendly	Shows tension release	Agrees	Asks for suggestion	Asks for opinion	Asks for information	Gives information	Gives opinion	Gives suggestion	Disagrees	Shows tension	Seems unfriendly
Twitter #LCSM and #LungCancer	'I'm so saddened & disappointed. My condolences to her loving family & the lung cancer community' 'Thanks everyone who walked today'	'US #1 cancer killer is Lung Cancer: 433 die daily,160 K die yearly! Lungs are sexy, too!'	'Agreed! Either light up the White House for every cancer or don't light it up at all! #WhyIsThisPink'	'How might doctors and other healthcare providers better integrate #yogatherapy into their offerings?'	'Which lung cancer group(s) do you favour? Why do you like them?'	'Which Genetic Tests Are Important for Metastatic Lung Cancer Treatment?'	'#BreakingNews: FDA approves new dosing option for ALUNBRIG (brigatinib) in the US'	'Let's not forget that #lungcancer takes more lives than the next four cancers put together. More research clearly needed'	'For those who've lost someone to lung cancer. Join me at the Life & Breath Rally.'	'This seems wrong'	'Damn it! xxx died' 'This f*kin' sucks. She so wanted to be in Oahu with her family. I hope she was.'	'… there is no need to be rude. I am a cancer survivor. I don't think I brought it on myself & I changed my career to advocate for those who feel their voices are lost…'
Facebook Lung Cancer Support Group	'Sending love and prayers' 'Good luck next week'	'Nobody is getting out of here alive. Life is a 100% fatal disease.'	'I agree with you…..hospitals are scary' 'I totally agree. Family and friends may be supportive but they don't understand what we're going through and feeling like everyone on here'	'Where and how do all of you find your inner strength? I'm so afraid. Please share what helps you.'	'Is there hope for Lung cancer? Need some warriors here to post some good news for a change….'	'Does it mean that it's not advanced NSCLC since my dad tested negative for the PD-L1?'	'Keytruda did not work for me because I have a ALK mutation in my cancer. I went to different oncologist and she put me on a targeted therapy for that mutation and my tumours are shrinking.'	'It's awful but it gets better when the radiation stops. Just work hard on staying hydrated and eating what you can.'	'Take it day by day and test by test….Stay away from the internet searches and Dr. Google.'		'we need to STOP the Stigma and not worry about who smokes, smoked or never smoked…it causes so much upset and hard feelings because so many people still 'assume' that you had to smoke to get Lung Cancer'	'SPAM!…You are in violation of the GROUP RULES!' 'Comments like that are so ignorant'
Macmillan.org.uk lung cancer discussion forum	'I'm so so sorry for your loss. Your mom fought and lost but she is now in a better place. I don't know what to say other than I'm thinking of you and her. Please take care and your time to grieve.'	'My dad's taste has completely changed! He's like a pregnant woman!'	'My grandad has been having the sweats! Definitely agree. They should make these symptoms more aware. Especially considering how aggressive this cancer is.'	'My dad has advanced lung cancer. …. Does anyone have any suggestions for good support pillows that helped them or their loved ones?	'If the 4 chemo sessions I'm going to get are only going to give me a short time extra and in that short time I'm going to feel dreadful, then what's the point? Does that sound a bit defeatist?'	'Found out today that if the disease progresses they would consider second line treatments such as Docetaxel or +/− Neratinib. Has anyone any experience of these?'	'I am brand new and so thankful that I found your site. I went to the ER yesterday for abdominal pain and they did an abdominal CT scan. They found 2 small nodules (2 mm and 3 mm) in one of my lungs. It says 2/3, stable densities suggesting benignity.'	'Things will get easier when you have a treatment plan and can get your head round it and see the way forward.'	'You can carry out a search within the Lung group typing some words into the search bar and it should list all discussions containing that text.' 'Have you been allocated a Macmillan nurse? If not speak to the oncology department.'		'The consultant didn't want to put a time frame on her life expectancy … how the hell do you get through this? How do I explain to my 4 year old when the time comes that grandma isn't coming over again?'	'My original post was supposed to provide my Dad with some positive stories and tips…I dont think i can share this thread with him after the last post. Im sympathetic to your situation but just ask you be the same to mine'

Table 4. Frequency of the four different functions of social support

Source	Total number of posts analysed	Emotional support	Instrumental support	Informational support	Companionship support	Not considered social support
Twitter #LCSM and #LungCancer	2897	168 (5.8%)	9 (0.3%)	842 (29.1%)	686 (23.7%)	1307 (45.1%)
Facebook Lung Cancer Support Group	2597	1333 (51.3%)	0 (0%)	1421 (54.7%)	71 (2.7%)	0 (0%)
Macmillan.org.uk lung cancer discussion forum	266	178 (66.9%)	0 (0%)	174 (65.4%)	33 (12.4%)	0 (0%)

The categories were not considered mutually exclusive

all of you. Love my tribe. Let's do this', 'Anyone can do something…A handful of us started #LCSM' and 'Join us on November 2 in Washington DC'. The higher percentage of companionship support posts on Twitter, may be attributed to the use of the #LCSM as a means of forming topic communities where large groups of users, who do not need to be connected through existing 'follower' networks, can interact within the constraints of Twitter's digital architecture.[48] The #LCSM hashtag was created with the intention to 'unite patients, caregivers, advocates, healthcare providers and researchers to discuss ways to improve lung cancer diagnosis, treatment, research, patient outcomes, caregiving, information sharing and public support'. It connects those participating in the pre-scheduled online discussion and is supported by posts such as 'If anyone is just lurking tonight, please blank tweet the hashtag #LCSM so we know you're out there. We're a friendly bunch' and 'Remember to include #lcsm in your tweets…'. This suggests that through the use of hashtags and followers[49] Twitter can provide a source of social and community support to people affected by lung cancer in knowing that they are not alone and in building self-esteem, confidence and social validation.[50]

Instrumental support was not present in any of the posts in the Lung Cancer Support Group on Facebook nor the lung cancer discussion forum on Macmillan.org.uk. It was however present on Twitter (0.3%) as users, often from the United States, requested financial support to fund their treatment 'Help me complete #LungCancer treatment #Donate #crowdfund Please retweet!'

Other topics that are notably scarce in our samples include stigmatization and trolling.

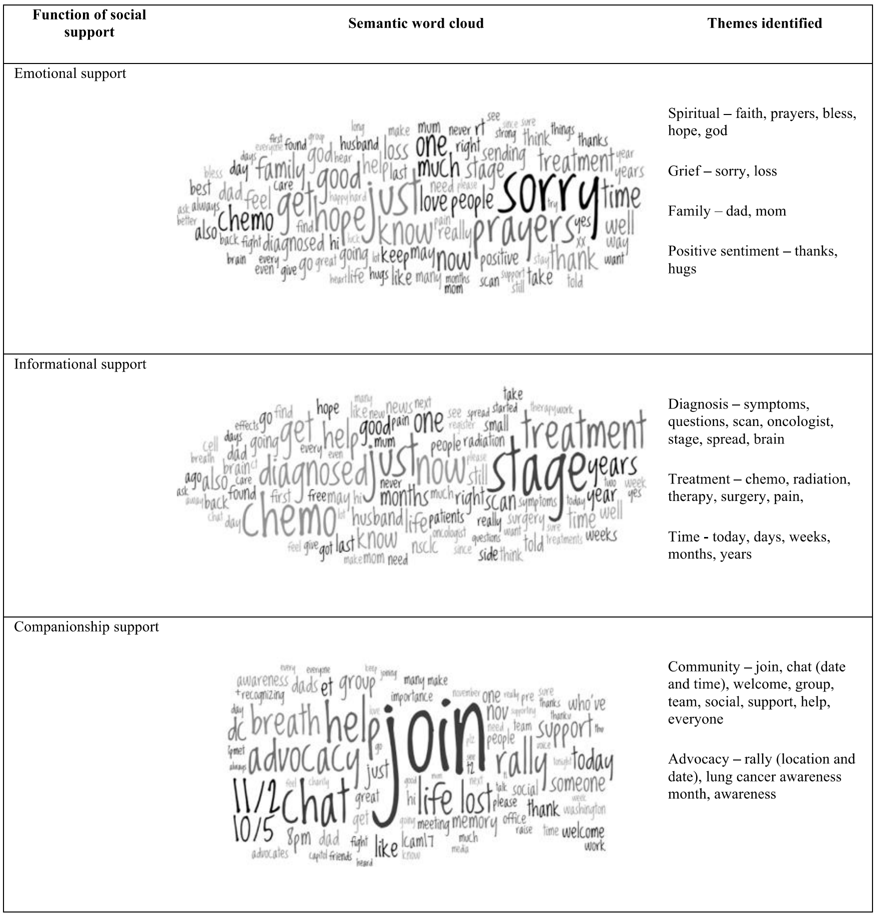
Although health-related stigmatization on social media has been reported in other research[51] and in specific about lung cancer,[52] here it was seen in only 1.8% of Twitter hashtags, 0.5% of posts to the Lung Cancer Support Group on Facebook and not at all in the lung cancer discussion forum on Macmillan.org.uk (0%). Examples showing how it was manifested include: 'A lot of people with cancer are afraid to talk about it, especially Lung Cancer, since some will just assume you smoke and you did it to yourself' (Facebook user), 'Lung cancer is the biggest killer yet there is no education on it. I do believe it is the stigma of smoking that is associated with it…it is the first thing people say to me' (Facebook user) and 'We're fighting stigma that holds lung cancer back from broad public sympathy despite being biggest cancer killer' (Twitter user).

Trolling is defined as 'the practice of behaving in a deceptive, destructive, or disruptive manner in a social setting on the internet with no apparent instrumental purpose'.[53] Unfriendly social interactions were represented by <1% of posts in the corpus of data from the three social media platforms; contrasting with a recent analysis showing that 24% of online trolling incidents are associated with health related topics.[54]

Limitations in our study reflect the self-selective nature of contributors, as a social media account is required in order to author a post. Given the 'digital health divide'[55] and recent statistics suggesting that most people diagnosed with lung cancer are 65 years or older,[56] the sample does not perfectly represent the population of people diagnosed and living with this condition. Analyses of social media data has its own limitations due to the presence of misinformation and a lack of provenance of account holders,[23] a topic that has recently attracted media attention due to use of bot factories to boost followers and the creation of 'fake news'.[57] In addition, the cross sectional nature of the study did not take into account variations in activity that may have occurred over time, which can be useful for examining evolving narratives during cancer progression.[49] Althoughthe data from the Lung Cancer Support Group on Twitter and discussion forum on Macmillan.org.uk were extracted manually, the Symplur Transcript and Analytics tool was used to extract the Twitter data, making the sampling less easy to verify. Other automated social media mining techniques and natural language processing tools are available; however, in some cases these require software and data access licenses, and can vary in their accuracy and effectiveness due to the scope and quality of data available and the types of social media for which they are suited.

In conclusion, our findings, based on a systematic analysis of comparable lung cancer posts on three social media platforms, indicate that although all three are being used to disseminate information about lung cancer, the Lung Cancer Support Group on Facebook and lung cancer discussion forum on Macmillan.org.uk, by virtue of their digital architecture, user-base and self-moderating communities, are more successful in their utility for social interaction and emotional and informational social support. While the sample derived from Twitter hashtags contained fewer posts related to social support across the four categories, posts tagged #LCSM showed the greatest degree of companionship support, revealing how the affordances of this platform can be shaped by its users through the use of a community hashtag. Further analysis also revealed an unanticipated sub-category of spiritual support, which featured uniquely in the Lung Cancer Support Group on Facebook and warrant additional research, as well as limited evidence of resentment about the comparative stigmatization of lung cancer compared with other types of cancer.[10]

These findings provide tentative insights into the social and supportive value of different social media, and show how interactions may be shaped both by the configuration and moderation of the platforms and by users self-organizing around groups or hashtags. They suggest that healthcare providers and policy makers wishing to provide supportive interventions via social media, including the use of social media to reduce stigma,[58] should prioritize community-based forums over general social media broadcasting. Likewise, academics and public health analysts wishing to study lung cancer via social media should carefully consider the types of data likely to appear on different platforms and its suitability for answering their research questions (e.g., whether through depth or volume). Importantly, the results also provide empirical evidence that people affected by lung cancer, and those supporting them (e.g., healthcare teams, family,



Function of social support	Semantic word cloud	Themes identified
Emotional support		Spiritual – faith, prayers, bless, hope, god Grief – sorry, loss Family – dad, mom Positive sentiment – thanks, hugs
Informational support		Diagnosis – symptoms, questions, scan, oncologist, stage, spread, brain Treatment – chemo, radiation, therapy, surgery, pain, Time - today, days, weeks, months, years
Companionship support		Community – join, chat (date and time), welcome, group, team, social, support, help, everyone Advocacy – rally (location and date), lung cancer awareness month, awareness

Fig. 3 Semantic word clouds visualizing the frequency of words by social support function

carers, clergy), should consider online communities as an additional source of social support during times of crisis.

Although hand searching and content analysis has proven effective in identifying and comparing the key types and expressions of social support for lung cancer manifested on these social media platforms further research is needed to unpick, replicate and extend these findings with larger samples of data. Automated data mining and natural language processing techniques enable the capture and analysis of much larger volumes of data across multiple social media platforms, offering potential to create a greater degree of precision[59] when combined with appropriate qualitative analysis. Nevertheless, it is essential that third party users, including healthcare providers, acknowledge the sensitivity of users' data, albeit it has been voluntarily placed in the public domain, since current research ethics guidelines governing use of aggregated social media data remain inconsistent.[60]

METHODS

Study design

For the purposes of this study we chose three social media platforms, which have different characteristics. Twitter and Facebook are general platforms for information sharing and social networking. In the case of Twitter, communities and topics are often collectively organized around a set of hashtags, which can then be searched for to understand particular issues, as in this study. Users of Facebook can set up discussion groups focused on certain topics, including lung cancer. Specialist organizations, such as Macmillan Cancer Support, may also set up condition-specific online discussion forums, where users can contribute content and others can comment. Users of social media must create an online profile and account ID when registering for these platforms. In doing so, it is within their gift to disclose personal information if they so choose. This variability results in some users being more easily identifiable as a patient, carer or family member than others. Given this incompleteness of profile data, our analysis treats patients and their carers or family members as one group of people affected by lung cancer.

To compare lung cancer-related interactions on different social media we took a multi-stage approach: (a) extracting and screening posts appearing in each social media platform, (b) classifying posts using Bales' IPA,[28] (c) categorizing posts according to the four functions of social support and (d) analyzing the 100 most frequent keywords to generate semantic 'word clouds' (using Wordle[61]) to visualize the frequency of terms used in posts associated with each form of social support.

Data sources and screening

Twitter: a sample of 3000 tweets was extracted using Symplur's Transcript and Analytics tool in December 2017, using the #LungCancer and #LCSM hashtags.[62] These hashtags were selected due to their specific relevance to the condition and as a means of reducing the amount of 'noise' presented in the data. The sample of Twitter data extracted and analysed in this study does not encompass the universe of all Twitter data available. Due to the limitations of the Transcript and Analytics tool, the sample size is limited to 1500 tweets per hashtag and the time frame for the extracted tweets was between 11:55 p.m. on 30 September 2017 and 01:00 a.m. on 1 November 2017. Although the transparency of Symplur's search and sampling algorithms has been criticized, its use in over 280 published research articles[63] provided justification for its use in this study. The data extracted included the Twitter account ID and the text in the tweet. Retweets and tweets that were not in English were excluded from the categorization stage. Url links and images included in the tweets were not captured or reviewed during the screening.

Facebook: using the Facebook search functionality, we searched for lung cancer and in doing so identified the largest lung cancer community available on Facebook. Known as the Lung Cancer Support Group, this community was established in 2015 and is a closed group for lung cancer patients, survivors, caregivers and loved ones. Closed groups are members only groups, where the group's existence is visible to anyone with a Facebook account. Membership of the group, however, is granted through the group administrators. Access to the Lung Cancer Support Group was granted via correspondence sent to the group administrators explaining the intent behind our request. As of 30 December 2017, it had 7975 members and on this same date all wall posts and replies that were posted between 1 and 31 October 2017 were identified by using the search functionality and stipulating the posts could be posted by 'anyone', 'anywhere' and in 'October 2017'. A sample of the original post and any associated replies, along with the author and the date of post were manually extracted for further analysis. Posts that were not in English were excluded from the categorization stage. Url links and images included in the posts were not captured or reviewed during the screening.

Macmillan.org.uk**:** Macmillan Cancer Support is a UK based charity that was founded in 1911 and provides specialist healthcare, information, financial and emotional support to people affected by cancer. It has an online community of over 100,000 members[64] spanning all forms of the condition. The lung cancer discussion forum on Macmillan.org.uk was identified through the search functionality of the online community homepage and the list of discussion threads was then filtered based on the start date 1 October 2017. All original posts and replies posted between 1 and 31 October 2017 were identified and manually extracted. The total sample of 266 posts was included in the categorization stage.

Categorization of posts by the application of Bales' IPA and the social support taxonomy

The same sample of lung cancer posts included in the categorization stage were reviewed against the 12 categories of group interaction defined in Bales' IPA (Fig. 2) as well as the four functions of social support (i.e., emotional, informational, instrumental, companionship). Each post was considered a single unit of interaction and the categories were not considered mutually exclusive when applied to the sample of posts.

Analysis of keywords

The content of lung cancer posts meeting each category of social support was analysed by inputting the data into the text visualization tool Wordle.[61] The resulting semantic word clouds map the frequency and co-occurrence of different terms appearing in a corpus of text and can be used by researchers to compare the topics and sentiment appearing in different text.[65] The word clouds produced in this study represent the top 100 words with the highest frequency of occurrence, in alphabetical order. Hashtags and terms such as 'lung', 'cancer' and 'lungcancer' were removed, as these had already been used to screen the social media posts for inclusion in the study.

Ethical considerations and informed consent

Although the data available on Twitter, public Facebook pages and the Macmillan.org.uk discussion forum exist in the public domain and can therefore be mined for research purposes without the need to obtain explicit informed consent from the data subjects,[66] ethical research conduct and digital etiquette are nevertheless required. As recommended by a recent review on the readiness of ethics guidelines to address research involving the secondary use of social media data,[60] we applied relevant sections from guidance developed by the UK Economic and Social Research Council,[67] the British Psychological Society[68] and the Association of Internet Researchers.[69] In the case of the closed Facebook group, access and agreement to extract and analyse the posts was requested from the group administrators. In order to protect the anonymity of the post authors, their account ID and any reference to other account IDs (e.g., the use of @) were removed). Ethics approval was provided by the University of Edinburgh's Institutional Review Board.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary.

DATA AVAILABILITY

The data used in this study is publicly available on Twitter, Facebook and Macmillan.org.uk.

ACKNOWLEDGEMENTS

C.P. is a collaborator on three RCUK-funded research programmes associated with this study: The Administrative Data Research Centre for Scotland, ESRC, Grant number ES/L007487/1; The Science and Practice of Social Machines EPSRC, EP/ J017728/1 and the Farr Institute for Health Informatics Research (Scotland) MRC, Grant number MR/K007017/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Symplur for the use of their Transcript and Analytics tool, users of Twitter, the Lung Cancer Support Group on Facebook and the lung cancer discussion forum on Macmillan.org.uk for posting about their experiences with lung cancer and for the administrators of the Facebook group for granting access to this closed online community.

AUTHOR CONTRIBUTIONS

J.T. and C.P. conceived and planned the study. J.T. extracted, screened and analysed the data, with input from C.P. J.T. first-drafted the manuscript and C.P. edited and contributed to the writing. Both J.T. and C.P. provided final approval of the completed manuscript and are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved. The authors received no specific funding for this work.

ADDITIONAL INFORMATION

Supplementary Information accompanies the paper on the *npj Digital Medicine* website (https://doi.org/10.1038/s41746-019-0124-y).

Competing interests: J.T. is a self-funded PhD student at the University of Edinburgh, supervised by C.P. She also works as a management consultant for Ernst and Young Switzerland although her employers were not involved in this research. C.P. declares no competing interests.

REFERENCES

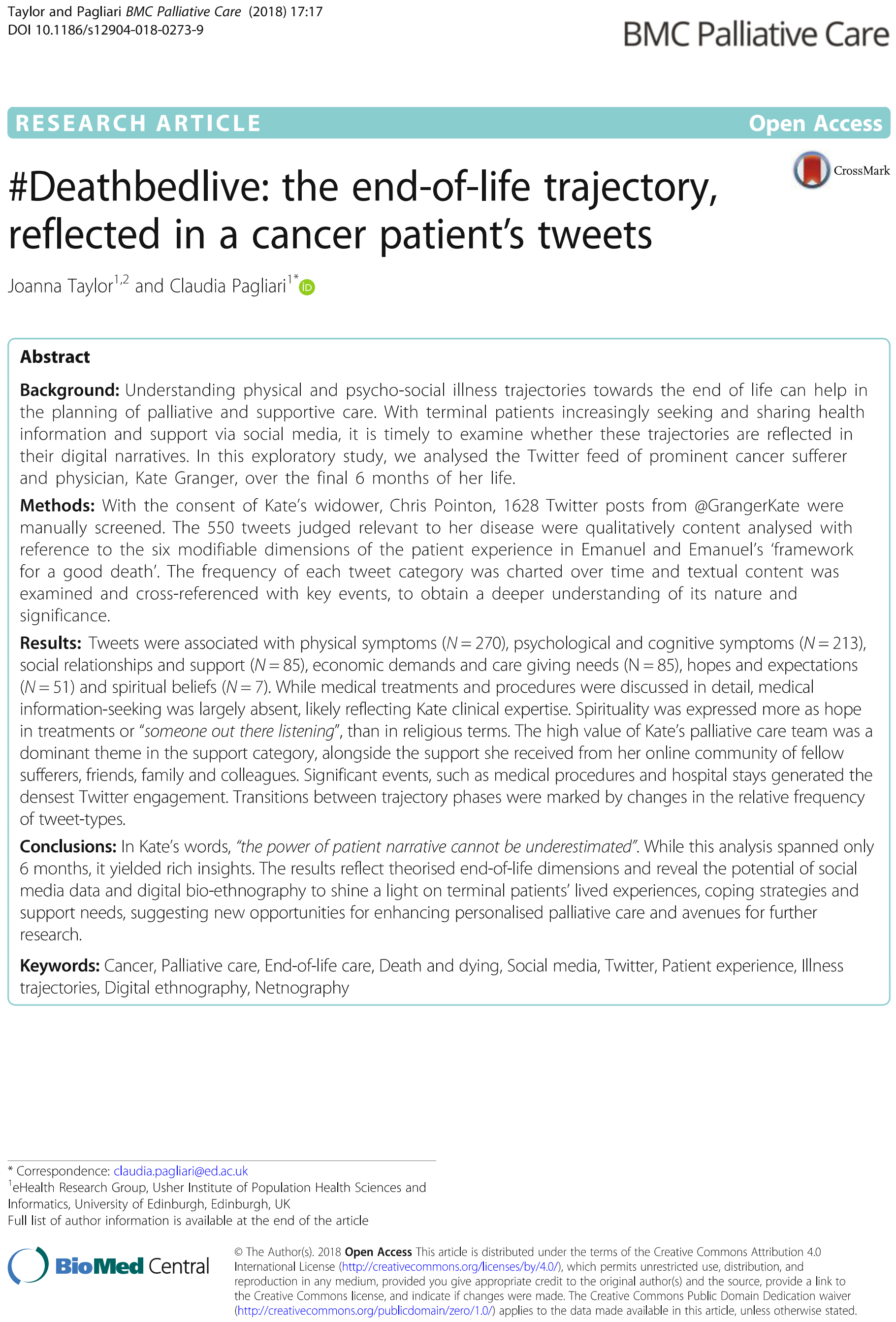
1. WHO. *Cancer Fact Sheet*. http://www.who.int/mediacentre/factsheets/fs297/en/ (2018).
2. lungcancer.org. *Types and Staging of Lung Cancer*. https://www.lungcancer.org/find_information/publications/163-lung_cancer_101/268-types_and_staging (2018).
3. Macmillan.org.uk. *A practical guide to understanding cancer - Lung Cancer*. http://be.macmillan.org.uk/Downloads/CancerInformation/CancerTypes/Mac11632LungE16Low-res-PDF20180401HS.pdf (2017).
4. Samet, J. M. et al. Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clin. Cancer Res.* **15**, 5626–5645 (2009).
5. Brown Johnson, C. G., Brodsky, J. L. & Cataldo, J. K. Lung cancer stigma, anxiety, depression, and quality of life. *J. Psychosoc. Oncol.* **32**, 59–73 (2014).
6. Cataldo, J. K. & Brodsky, J. L. Lung cancer stigma, anxiety, depression and symptom severity. *Oncology* **85**, 33–40 (2013).
7. Chapple, A., Ziebland, S. & McPherson, A. Stigma, shame, and blame experienced by patients with lung cancer: qualitative study. *BMJ* **328**, 1470 (2004).
8. Marlow, L. A. V., Waller, J. & Wardle, J. Does lung cancer attract greater stigma than other cancer types? *Lung Cancer* **88**, 104–107 (2015).
9. Sriram, N. et al. Attitudes and stereotypes in lung cancer versus breast cancer. *PLoS ONE* **10**, e0145715 (2015).
10. Else-Quest, N. M., LoConte, N. K., Schiller, J. H. & Hyde, J. S. Perceived stigma, self-blame, and adjustment among lung, breast and prostate cancer patients. *Psychol. Health* **24**, 949–964 (2009).
11. Bresnahan Mary, J., Silk, K. & Zhuang, J. You did this to yourself! Stigma and blame in lung cancer. *J. Appl Soc. Psychol.* **43**, E132–E140 (2013).
12. Eldridge, L. *Lung Cancer - The bias, the stigma, the shame, the blame and guilt*. https://www.lungcancerfoundation.org/2013/06/lung-cancer-the-bias-the-stigma-the-shame-blame-and-guilt/ (2013).
13. Kaplan, A. A. H. M. Users of the world, unite! The challenges and opportunities of Social Media. *Bus. Horiz.* **53**, 59–68 (2010).
14. Statista. *Number of social network users worldwide from 2010 to 2021 (in billions)*. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ (2018).
15. Hamm, M. P. et al. Social media use among patients and caregivers: a scoping review. *BMJ Open* **3** (2013).
16. Patel, R., Chang, T., Greysen, S. R. & Chopra, V. Social media use in chronic disease: a systematic review and novel taxonomy. *Am. J. Med* **128**, 1335–1350 (2015).
17. Koskan, A. et al. Use and taxonomy of social media in cancer-related research: a systematic review. *Am. J. Public Health* **104**, e20–e37 (2014).
18. Gualtieri, L. & Akhtar, F. Y. Cancer patient blogs: how patients, clinicians, and researchers learn from rich narratives of illness. in *Information Technology Interfaces (ITI), Proceedings of the ITI 2013 35th International Conference*. 3–8 (IEEE, 2013). https://ieeexplore.ieee.org/document/6648988.
19. Tsuya, A., Sugawara, Y., Tanaka, A. & Narimatsu, H. Do cancer patients tweet? Examining the Twitter use of cancer patients in Japan. *J. Med Internet Res* **16**, e137 (2014).
20. Crannell, W. C., Clark, E., Jones, C., James, T. A. & Moore, J. A pattern-matched Twitter analysis of US cancer-patient sentiments. *J. Surg. Res.* **206**, 536–542 (2016).
21. Sutton, J. et al. Lung cancer messages on Twitter: content analysis and evaluation. *J. Am. Coll. Radiol.* **15**, 210–217 (2017).
22. Capurro, D. et al. The use of social networking sites for public health practice and research: a systematic review. *J. Med Internet Res* **16**, e79 (2014).
23. Moorhead, S. A. et al. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J. Med Internet Res* **15**, e85 (2013).
24. Taylor, J., Osborne, M. & Pagliari, C. *Understanding the social dynamics of Twitter, Facebook and Diabetes.co.uk and their value implications for patients and health researchers*. https://www.thinkmind.org/download.php?articleid=global_health_2015_5_10 (2015).
25. *Twitter*. http://www.twitter.com.
26. *Facebook - Lung Cancer Support Group*. https://www.facebook.com/groups/733948110039512/?fref=nf
27. Macmillan.org.uk. *Macmillan.org.uk - Lung Cancer Discussion Forum*. https://community.macmillan.org.uk/cancer_types/lung-cancer/ (2018).
28. Bales, R. F. *Interaction process analysis: a method for the study of small groups*. (Addison-Wesley Press, Cambridge, MA, 1951).
29. Langford Catherine Penny, H., Bowsher, J., Maloney Joseph, P. & Lillis Patricia, P. Social support: a conceptual analysis. *J. Adv. Nurs.* **25**, 95–100 (2008).
30. Wills, T. A. Social support and interpersonal relationships. *Rev. Personal. Soc. Psychol.* **12**, 265–289 (1991).
31. Uchino, B. *Social Support and Physical Health: Understanding the Health Consequences of Relationships* (Yale Univ. Press, 2004).
32. Bossetta, M. The digital architectures of social media: comparing political campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *J. Mass Commun. Q.* **95**, 471–496 (2018).
33. Kamps, H. J. *Who are Twitter's verified users?* https://medium.com/@Haje/who-are-twitter-s-verified-users-af976fc1b032 (2015).
34. Abdulla, S. *What percentage of Twitter users are fake?* https://www.quora.com/What-percentage-of-Twitter-users-are-fake (2017).
35. Wise, K., Hamman, B. & Thorson, K. Moderation, response rate, and message interactivity: features of online communities and their effects on intent to participate. *J. Comput.-Mediat. Commun.* **12**, 24–41 (2006).
36. Balatsoukas, P., Kennedy, C. M., Buchan, I., Powell, J. & Ainsworth, J. The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness. *J. Med Internet Res* **17**, e141 (2015).
37. #LCSM. *#LCSM Chat Topic 10/5: Lung Cancer Advocacy: How we can help ourselves*. https://lcsmchat.com/2017/10/04/lcsm-chat-topic-105-lung-cancer-advocacy-how-we-can-help-ourselves/ (2017).
38. Wiesenhofer, H., Ebner, M. & Kamrat, I. Is Twitter an individual mass communication medium? in *Society for Information Technology & Teacher Education International Conference* (AACE, Chesapeake, VA, 2010). https://www.learntechlib.org/p/33604/.
39. Macskassy, S. A. On the Study of Social Interactions in Twitter. in *Proc. 6th International AAAI Conference on Weblogs and Social Media*. 226–233 (ICWSM, 2012).
40. Kaplan, B. H., Cassel, J. C. & Gore, S. Social support and health. *Med. Care* **15**, 47–58 (1977).
41. Emanuel, E. J. & Emanuel, L. L. The promise of a good death. *Lancet (Lond., Engl.)* **351**(Suppl 2), SII21 (1998).
42. Murray, S. A. et al. Archetypal trajectories of social, psychological, and spiritual wellbeing and distress in family care givers of patients with lung cancer: secondary analysis of serial qualitative interviews. *BMJ* **340**, c2581 (2010).
43. Murray, S. A. et al. Patterns of social, psychological, and spiritual decline toward the end of life in lung cancer and heart failure. *J. Pain. Symptom Manag.* **34**, 393–402 (2007).
44. Scott, A. M., Marilyn, K., Kirsty, B., Allison, W. & Benton, T. F. Exploring the spiritual needs of people dying of lung cancer or heart failure: a prospective qualitative interview study of patients and their carers. *Palliat. Med.* **18**, 39–45 (2004).
45. Holmberg, K., Bastubacka, J. & Thelwall, M. "@God please open your fridge!" Twitter messages to @God in content analysis: Hopes, humor, spirituality, and profanities. *J. Relig. Media Digit. Cult.* **5**, 339–355 (2016).
46. Lundby, K. Patterns of belonging in online/offline interfaces of religion. *Inf., Commun. Soc.* **14**, 1219–1235 (2011).
47. Campbell, H. A. Understanding the Relationship between Religion Online and Offline in a Networked Society. *J. Am. Acad. Relig.* **80**, 64–93 (2012).
48. Bruns, A. & Burgess, J. E. The use of Twitter hashtags in the formation of ad hoc publics. In *Proc. 6th European Consortium for Political Research (ECPR) General Conference* (QUT ePrints, 2011). https://eprints.qut.edu.au/46515/.
49. Taylor, J. & Pagliari, C. #Deathbedlive: the end-of-life trajectory, reflected in a cancer patient's tweets. *BMC Palliat. Care* **17**, 17 (2018).
50. Ko, H.-C., Wang, L.-L. & Xu, Y.-T. Understanding the different types of social support offered by audience to A-list diary-like and informative bloggers. *Cyber. Behav. Soc. Netw.* **16**, 194–199 (2013).
51. Boudewyns, V., Himelboim, I., Hansen, D. L. & Southwell, B. G. Stigma's effect on social interaction and social media activity. *J. Health Commun.* **20**, 1337–1345 (2015).
52. Occhipinti, S. et al. Lung cancer stigma across the social network: patient and caregiver perspectives. *J. Thorac. Oncol.* **13**, 1443–1453 (2018).
53. Buckels, E. E., Trapnell, P. D. & Paulhus, D. L. Trolls just want to have fun. *Personal. Individ. Differ.* **67**, 97–102 (2014).
54. Statista. *On which of the following topics have you seen trolling behavior on the internet?* https://www.statista.com/statistics/380051/topics-witness-trolling-behavior-internet/ (2018).
55. Hall, A. K., Bernhardt, J. M., Dodd, V. & Vollrath, M. W. The digital health divide: evaluating online health information access and use among older adults. *Health Educ. Behav.* **42**, 202–209 (2015).
56. Society, A. C. *Key Statistics for Lung Cancer*. https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html (2018).

57. Confessore, N., Dance, G. J. X., Harris, R. & Hansen, M. *The Follower Factory - Everyone wants to be popular online. Some even pay for it. Inside social media's black market*. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html (2018).
58. Carter-Harris, L. ES07.04 leveraging social media to change the public conversation on lung cancer stigma. *J. Thorac. Oncol.* **13**, S202 (2018).
59. Munoz-Garcia, O. & Navarro, C. Comparing user generated content published in different social media sources. 1–8 In *LREC Conference* (LREC, 2012). http://www.lrec-conf.org/proceedings/lrec2012/workshops/21.LREC2012%20NLP4UGC%20Proceedings.pdf#page=6.
60. Taylor, J. & Pagliari, C. Mining social media data: How are research sponsors and researchers addressing the ethical challenges? *Research Ethics*. **14**, 1–39, (2017).
61. Feinberg, J. *Wordle*. http://www.wordle.net (2014).
62. Symplur. *#LungCancer healthcare social media hashtag*. https://www.symplur.com/healthcare-hashtags/lungcancer/ (2018).
63. Symplur. *Healthcare Social Media Research*. https://www.symplur.com/healthcare-social-media-research/ (2018).
64. Macmillan.org.uk. *Online Community News Blog: 100,000 members!* https://community.macmillan.org.uk/blogs/b/community_news/archive/2016/02/10/100-000-members (2016).
65. McNaught, C. & Lam, P. Using Wordle as a supplementary research tool. *Qual. Rep.* **15**, 630+ (2010).
66. Hookway, N. 'Entering the blogosphere': some strategies for using blogs in social research. *Qual. Res.* **8**, 91–113 (2008).
67. ESRC. *Framework for research ethics*. http://www.esrc.ac.uk/_images/framework-for-research-ethics_tcm8-33470.pdf (2015).
68. BPS. *Conducting Research on the Internet - Guidelines for ethical practice in psychological research online*. http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf (2012).
69. Markham, A. & Buchanan, E. *Ethical Decision-Making and Internet Research*. http://www.aoir.org/reports/ethics2.pdf (2012).

A.2.4 BMC Palliative Care (published January 2018)

Altmetric attention score: 72 (as of 2 May 2020)

Number of citations: 8 (as of 2 May 2020)



Taylor and Pagliari *BMC Palliative Care* (2018) 17:17
DOI 10.1186/s12904-018-0273-9

BMC Palliative Care

RESEARCH ARTICLE Open Access

#Deathbedlive: the end-of-life trajectory, reflected in a cancer patient's tweets

Joanna Taylor[1,2] and Claudia Pagliari[1*]

Abstract

Background: Understanding physical and psycho-social illness trajectories towards the end of life can help in the planning of palliative and supportive care. With terminal patients increasingly seeking and sharing health information and support via social media, it is timely to examine whether these trajectories are reflected in their digital narratives. In this exploratory study, we analysed the Twitter feed of prominent cancer sufferer and physician, Kate Granger, over the final 6 months of her life.

Methods: With the consent of Kate's widower, Chris Pointon, 1628 Twitter posts from @GrangerKate were manually screened. The 550 tweets judged relevant to her disease were qualitatively content analysed with reference to the six modifiable dimensions of the patient experience in Emanuel and Emanuel's 'framework for a good death'. The frequency of each tweet category was charted over time and textual content was examined and cross-referenced with key events, to obtain a deeper understanding of its nature and significance.

Results: Tweets were associated with physical symptoms ($N = 270$), psychological and cognitive symptoms ($N = 213$), social relationships and support ($N = 85$), economic demands and care giving needs ($N = 85$), hopes and expectations ($N = 51$) and spiritual beliefs ($N = 7$). While medical treatments and procedures were discussed in detail, medical information-seeking was largely absent, likely reflecting Kate clinical expertise. Spirituality was expressed more as hope in treatments or "*someone out there listening*", than in religious terms. The high value of Kate's palliative care team was a dominant theme in the support category, alongside the support she received from her online community of fellow sufferers, friends, family and colleagues. Significant events, such as medical procedures and hospital stays generated the densest Twitter engagement. Transitions between trajectory phases were marked by changes in the relative frequency of tweet-types.

Conclusions: In Kate's words, *"the power of patient narrative cannot be underestimated"*. While this analysis spanned only 6 months, it yielded rich insights. The results reflect theorised end-of-life dimensions and reveal the potential of social media data and digital bio-ethnography to shine a light on terminal patients' lived experiences, coping strategies and support needs, suggesting new opportunities for enhancing personalised palliative care and avenues for further research.

Keywords: Cancer, Palliative care, End-of-life care, Death and dying, Social media, Twitter, Patient experience, Illness trajectories, Digital ethnography, Netnography

---

* Correspondence: claudia.pagliari@ed.ac.uk
[1] eHealth Research Group, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
Full list of author information is available at the end of the article

Background

In 1968 Glaser and Strauss described the advancement towards death as having the elements of time and shape [1] giving rise to the concept of the illness trajectory. Originally developed to describe how physical aspects of a patient's disease unfold through the phases of pre-trajectory, trajectory onset, living with disease progression, downward phase and dying [2], the concept has since been expanded to include psychosocial aspects of the patient experience, including their response to their illness, the people around them and the interventions undergone [3]. Emanuel and Emanuel's 'framework for a good death' [4] has been particularly influential in helping clinicians to better anticipate the needs of patients during the progression of their illness and to shape palliative care services. The Framework articulates six 'modifiable dimensions' of the patient experience related to 1) physical symptoms, 2) psychological and cognitive symptoms, 3) social relationships and support, 4) economic demands and care giving needs, 5) hopes and expectations, and 6) spiritual and existential beliefs.

Most research into illness trajectories originates from the fields of public health and social sciences, drawing on studies using qualitative or mixed-methods, with data typically gathered from cohorts of patients through focus groups, surveys and interviews. This research has revealed different illness trajectories for different terminal conditions, with the cancer trajectory described as a steady progression over a period of weeks-and sometimes years, punctuated with the positive and negative effects of oncology treatment, weight loss, reduction in physical performance and the impaired ability to self-care during the last few months, as shown in Fig. 1 [3]. Cancers can also have unique trajectories, depending on issues such as prognosis, pain, disfigurement and response to treatment [5] while the same type of cancer progression may be experienced differently as a consequence of personal and social factors such as resilience and availability of emotional support [6].

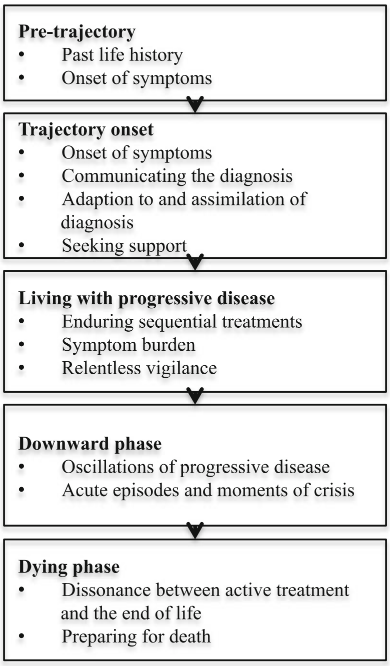


Fig. 1 Cancer illness trajectory phases [34]

Social media in health

Technological advances over the past decade have enabled widespread access to the Internet in most countries, with the use of social media sites becoming increasingly common. Social media are a class of online, often mobile, platforms that support the creation and exchange of user-generated content [7], a phenomenon sometimes referred to by the terms Web 2.0 or Social Web. Social media include generic platforms for networking, information sharing and collaboration (e.g. Facebook, Twitter, YouTube, LinkedIn), as well as online forums aimed at specific communities (e.g. PatientsLikeMe, Mumsnet). As health and wellness are dominant societal concerns, they account for a considerable amount of social media activity, and some analysts have referred to this as Health 2.0 or Medicine 2.0 [8].

Several uses of social media in research have been described in the literature. These include the deployment of social media platforms for the *conduct* of research, such as for online surveys [9], recruitment to studies [10] and participative 'citizen science' [11], and as a *source of data* for research [12]. The latter 'secondary uses' include social, business and policy research examining user demographics, interactions and networks, 'social listening' to understand sentiments associated with particular topics or products [13, 14] and 'big data' analytics to uncover new associations or predict future behaviours or outcomes. The term "infoveillance" has also been used to refer to the automated and continuous analysis of unstructured, free text information available on the Internet for the purpose of public health monitoring [15].

Social media use by cancer patients

A scoping review of studies involving social media use by patients and caregivers [16] reported that discussion forums, online support groups, social networking sites and micro-blogs, such as Twitter, dominated the published literature, with 11.3% of the 284 identified studies focusing on cancer. Examples included a study examining the use of Twitter and its role in the lives of cancer patients, which revealed that that the majority of tweets focused on psychological support [17], and a narrative synthesis of cancer patient blogs, which indicated that users share their experiences online in order to find information, maintain relationships with others, inform their health behaviour, and 'tell their story' about their treatment, diagnosis and symptoms and their experiences of health services [18].

Historically death and dying has been seen as a taboo topic for public discussion, however the willingness to talk openly about it online by terminally ill patients, their families and palliative health and social care professionals, has increased over recent years giving rise to an emerging field of research [19]. The first empirical study to have analysed data from Twitter for understanding disease was published in 2010 [20], yet among the many subsequent studies using Twitter in such research [21], we have been unable to find any empirical studies that have analysed how the online activity of cancer patients reflects the illness trajectories documented in previous research or explored the potential of this emerging data source to yield insights about cancer patients' experiences at the end-of-life.

Aims of this study

This study sought to systematically analyse the content of one prominent cancer patient's Twitter feed in the final 6 months of her life, in order to determine its fit with documented end-of-life trajectories and the 6 dimensions of the 'framework for a good death,' as well as to explore the value of social media data for understanding patients' personal experiences, life quality and coping strategies. The study was intended as exploratory and hypotheses-generating, with a view to providing insights to inform future research and the design of innovative palliative care services.

Kate granger

Kate Granger was an English geriatrician and campaigner for better patient care. In 2011 at the age of 29, she was diagnosed with a rare form of sarcoma, known as desmoplastic small-round-cell tumour, with a predicted to life expectancy of around 5 years. She was treated with P6 protocol chemotherapy and endured painful treatments, which she described in detail in her blogs "The Other Side and the Bright Side" [22]. Kate created her Twitter account (@GrangerKate) in March 2012 and, prior to her death on 23 July 2016, posted approximately 12,500 tweets and attracted approximately 48,000 followers [23].

Through her experiences as a patient, she and her husband founded the "#hellomynameis" campaign encouraging healthcare staff to introduce themselves to patients. They raised over £250,000 for local cancer charity the Yorkshire Cancer Centre Appeal and in 2015 she was awarded an MBE for her services to the British National Health Service (NHS). As a young woman familiar with social media, Kate's story represents a valuable opportunity to examine the emergence and progression of a personal narrative, in the public domain, about coping with terminal illness.

Methods

Although the cancer trajectory can last for years, and did so for Kate Granger, we chose to study the final 6 months of life, which represents the terminal phase and is commonly associated with preparation for death and the commencement of hospice services, for those fortunate enough to receive the latter [24]. Original tweets, re-tweets and responses posted between 1 January 2016 and 25 July 2016 using the account @GrangerKate, were manually extracted for categorization and analysis. The data extracted included the date and time of posting and the up to 140-character text contained within the tweet.

The tweets were then manually screened for their relevance to the disease, based on predefined inclusion and exclusion criteria, as described in Table 1.

Drawing on principles of digital ethnography [25] we used qualitative content analysis [26] to summarize, chart and interpret the eligible tweets. Tweet content was first categorised according to the six modifiable dimensions of the patient experience in the 'framework for a good death,' shown in Table 2, with each post treated

Table 1 Tweet inclusion and exclusion criteria

Inclusion criteria:
- Tweets posted by @GrangerKate's Twitter account
- Original tweets, annotated re-tweets and personal responses posted between 0:00:00 (UTC) 1 January 2016 and 23:59:59 (UTC) 25 July 2016
- Tweets that were considered directly relevant to the terminal condition

Exclusion criteria:
- Tweets posted on other Twitter accounts
- Simple re-tweets of other people's postings (with no further annotation)
- Tweets posted outside the timeframe indicated
- Tweets that were considered not directly relevant to the condition, such as those associated with the "#hellomynameis" campaign, Kate Granger's fundraising activities, her views on the NHS and politics as well as news media
- Images or URLs embedded within tweets

Table 2 Modifiable dimensions of the patient experience, from the 'framework for a good death' [4]

Modifiable dimension of the patient experience	Examples of specific concerns
Physical symptoms	Pain and fatigue
Psychological and cognitive symptoms	Depression, anxiety and confusion
Social relationships and support	Family, community, interests
Economic demands and care giving needs	Saving and income, personal care and nursing care
Hopes and expectations	Milestones and assessment of prognosis
Spiritual and existential beliefs	Religion, sense of purpose and meaning

as a single unit of interaction and the categories as non-exclusive. The narrative content of tweets was also examined, to obtain further contextual information about significant events and personal responses. The frequency of each category, as well as the occurrence of key events (such as medical procedures or transfer to hospice) were plotted over time and converged with the qualitative data in order to "learn the meanings, norms, patterns of a way of life" ([27]: pg13) and to enable comparisons to be made with published end-of-life trajectories. Images and web links included in the tweets were not reviewed during the screening.

Ethical considerations

Although the data available on Twitter exist in the public domain and can therefore be mined for research purposes without the need to obtain explicit informed consent from the data subjects [28], ethical research conduct and digital etiquette are nevertheless required. Given the lack of consensus on ethical principles for the secondary use of social media data, we applied relevant sections from guidance developed by the UK Economic and Social Research Council (ESRC) [29], the British Psychological Society (BPS) [30] and the Association of Internet Researchers (AoIR) [31], as recommended by a recent review on the readiness of ethics guidelines to address this type of research [32]. While the ERSC "Framework for Research Ethics" cautions that studies involving online respondents may involve more than minimal risk [29], the BPS "Guidelines for ethical practice in psychological research online" go further, by differentiating between participants who are identifiable or anonymous and those who are actively recruited or studied without their knowledge [30]. The AoIR guidance on "Ethical Decision-Making and Internet Research" provides a series of questions for researchers to consider when conducting studies of this type, and reflect the context, risks and management of the data and how the findings will be presented [31].

In the case of this study, written agreement to extract, analyse and publish the tweets posted by @GrangerKate was sought and obtained from Kate Granger's widower, Chris Pointon, via email.

Results

Kate's tweets were first screened for eligibility and those that were considered relevant to her condition were then classified and plotted on a visual timeline.

Tweet eligibility

Of the 1628 tweets posted by @GrangerKate during the 6-month period, 550 were considered relevant to her condition and therefore included in the classification, as described in Fig. 2. The remaining 1078 tweets were excluded for reasons such as their focus on her

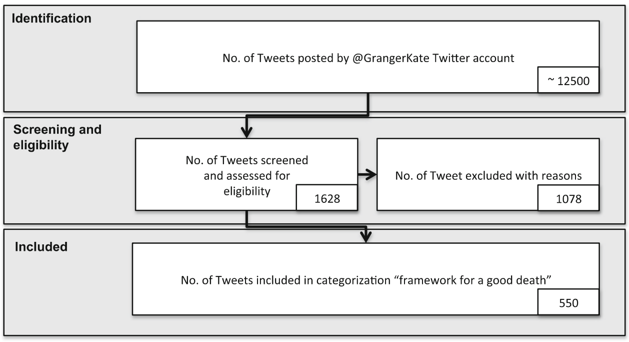


Fig. 2 Steps in screening and assessing the tweets

campaigning activities (388) as well as her views on the NHS (109) and politics (105), as summarised in Table 3.

The 550 tweets considered relevant to the condition, were classified according to the modifiable dimensions of the patient experience in Emanuel and Emanuel's 'framework for a good death'. The total number of tweets classified for each dimension is summarized in Table 4, including an example of the tweet narrative in each category.

Nature of tweets over time

Figure 3 visualises Kate Granger's digital end-of-life trajectory. The coloured lines represent the frequency of daily tweets, according to each of the six dimensions of the 'framework for a good death', plotted over the 6-month observation period. Similar to Barclay et al.'s study into the trajectories to death in residential care homes [33], the annotations in Fig. 3 describe key contextual events and the superscript shows the broad phases of the illness trajectory, which were evident in Kate's tweets.

Given that Kate had been diagnosed with cancer for 5 years, Corbin and Strauss' pre-trajectory and trajectory onset phases were not represented in the sample of tweets. However, Tweet characteristics indicative of the living with progressive disease, downward trajectory and dying phases were evident and these form the basis of our findings.

Living with progressive disease phase

The 'living with progressive disease' phase can in some cases last several years, where patients go through cycles of feeling and looking well and living life to the full, followed by periods of disease progression and illness. Although the full extent and duration of this phase was not analysed as part of this study, which focused on the final 6-months of life, initial findings indicate that Kate posted 600 tweets, from the beginning of the observation period on 1 January 2016 until 20 March 2016, of which only 16% were associated with her condition. She endured a series of sequential treatments and procedures including her seventh stent replacement in late January 2016, with the symptom burden being indicated by tweets such as "*started 2015 worrying about cancer. Start 2016 worrying about cancer*" and "*we've been living with the reality of cancer in our lives for nearly 5 years…it becomes the new normal*" – 1 January 2016.

Table 3 Number of tweets excluded and the reason for exclusion

Reason for exclusion	Number of tweets
Campaigns including "#hellomynameis" and fundraising activities	388
Professional work as a geriatrician	89
Hobbies including baking, flute and band practice	48
NHS in general	109
Politics including her support for the Junior Doctors' strike	105
Dissemination of news and research	38
Others, including general "thank you" messages and topics that could not be determined from the tweet narrative	437

Uncertainty seemed to pervade Kate's life during this phase as indicated by her tweet "*In pain. Horrible achy pelvic pain….Are you waking up cancer or is it in my head?*" – 21 February 2016. She was at home and still able to work, requiring minimal caregiver support and continued with her hobbies and campaigning activities. She oscillated between disease progression and treatment, with psychological and cognitive symptoms appearing to closely follow the physical symptoms. She had received support from a palliative care team from an early stage ("*Not been to hospice yet but had palliative care team support almost from day one…*" – 4 May 2016). Although treatment allowed some respite from symptoms of progressive disease ("*It kept me well for nearly 6 months. But was never going to cure me & was always a palliative measure…*" – 13 March 2016), living with the condition appeared exhausting in its relentlessness ("*I'm just so tired of being in pain. Nagging pain that I can't escape night after night. Had enough.*" – 21 April 2016).

Downward phase

Kate's downward phase was characterized by increasingly frequent episodes of illness and crisis over a 4-month period between 20 March 2016 and 21 June 2016, which is reflected in a marked increase in the frequency of her tweets. Peaks in this digital trail marked acute episodes, associated with treatments and procedures such as chemotherapy, radiotherapy and endoscopic retrograde cholangiopancreatographies (ERCP), and four stays in hospital over a 10-week period. By this stage of her illness she was no longer able to work and many of her tweets detailed the physical and psychological anxiety she experienced as well as the support provided by healthcare professionals and by her physical and virtual community of family, friends and followers. During this phase she posted 898 tweets of which 42% were associated with her condition.

During these acute episodes, tweets conveyed a lack of confidence in being able to cope ("*I'm exhausted of being 'brave', the expectation that I'm 'Dr Kate Granger' and therefore can cope with anything. I can't…*" – 3 May 2016) and acknowledging that the end was near ("*Perhaps I should just accept #deathbedlive is closer than I*

Table 4 Number of tweets per modifiable dimension

Modifiable dimension of patient experience	Number of tweets	Examples of tweets
Physical symptoms	270	"I am having a temporary nephrostomy this morning... With a view to then getting the radiotherapy done then performing a stent exchange" – 15 June 2016
Psychological and cognitive symptoms	213	"I'm not sure I can do this." – 3 April 2016
Social relationships and support	85	"Everybody being so lovely both in public & behind the scenes on Twitter. Thank you so much. Love our virtual family very much" – 8 May 2016
Economic demands and care giving needs	85	"After seeing my lovely palliative care nurse this a.m. we've decided hospice admission for symptom control best course of action." – 8 July 2016
Hopes and expectations	51	"Perhaps I should just accept #deathbedlive is closer than I hoped it was & get my final preparations finished." - 29 March 2016
Spiritual and existential beliefs	7	"A comfortable night, just one would be so appreciated if anyone is listening. Running on empty and a nonstop few days coming up." – 3 March 2016

hoped it was & get my final preparations finished." – 19 July 2016). Respite came from being at home ("*Very happy to be on the way home. Thank you so much to #nhs for scraping me up & putting me back together…"* – 18 May 2016) and the positive messages of thanks and support given by those around her ("*Aww having a little smiley weep at all the wonderful messages. Everyone is just so kind. Thank you so much xx #kateshomecoming*" – 25 June 2016).

Dying phase

In the 4-weeks, prior to her death on 23 July 2016, Kate experienced a rapid decline in health, which was mirrored by a decline in the number of tweets posted, down to 130 of which 57% were associated with her condition. Kate's acknowledgement of the terminal phase is indicated in her first reference to the transition to hospice care, with the tweet "*Hospice waiting list was going to be well into next week so decided to bite the bullet and come home…"* – 23 June 2016. She appeared to be preparing for death during the 4-week period after meeting with the hospice consultant on 29 June 2016, before being transferred to St Gemma's Hospice on 8th July. In contrast to the previous phases, only 30% of Kate's Tweets in the dying phase referred to her physical and psychological symptoms, and the remainder instead focused on the social and caregiving support she received from her palliative care and hospice team, and her family, friends and online followers.

Discussion

We have described the pattern of physical, psychological, social and care needs of high profile sarcoma patient Kate Granger, as represented in the posts from her

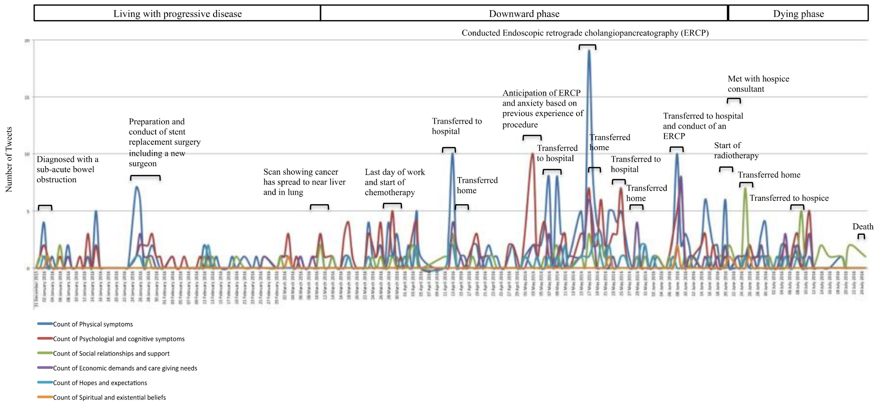


Fig. 3 Kate Granger's illness trajectory

Twitter account over the 6-month period prior to her death. Our analysis indicates that the digital manifestation of cancer progression maps to the illness trajectories described in existing palliative care research and to the key dimensions of the patient experience articulated in Emanuel and Emanuel's 'framework for a good death'. Our analysis has also tested the use of social media data and digital ethnographic approaches for exploring the lived experiences of patients nearing the end of life.

Parallels between the digital end-of-life trajectory and existing models

In interpreting the digital trajectory seen in Kate Granger's Twitter postings, we drew on the general cancer trajectory described in the literature, in which patients experience an onset of incurable disease and a short and rapid decline over a few months. This type of trajectory involves a reasonably predictable decline in physical health, with progression punctuated by the positive and negative effects of palliative oncology treatments [3]. While our analysis focuses on the later stages of the cancer illness trajectory it is important to consider the similarities and differences between the patterns seen in Kate Granger's tweets and these general trajectory descriptions. In Kate's case, similarities can be drawn in relation to her endurance of sequential treatments, the acute episodes and moments of crisis as well as her preparations for death. As noted in the introduction, the trajectory of different cancers varies, to some extent, and it is important to bear this in mind when seeking to interpret the patterns observed in individual cases. For example, Reed and Corner's research into the illness trajectory of metastatic breast cancer identified a "rollercoaster" trajectory, where the typical duration was 2-5 years and patients, similarly to Kate Granger, experience oscillations of disease progression, treatment and restoration of well-being, as well as acute crisis episodes [34].

Emanuel and Emanuel's 'framework for a good death' proved a useful tool for classifying the data, which showed a good fit with their six 'modifiable dimensions of the patient experience', and helped in the plotting of physical symptoms and psychosocial responses. For example, the differentiation of tweet types aided the interpretation of changes in Twitter activity between the different phases of the illness trajectory, such as the priority accorded to physical and medical issues in the earlier stages, compared to the focus on people and comfort-giving at the end.

It is interesting to note that, based on her tweets over the 6-month period, Kate made no explicit reference to spirituality, which contrasts with previous observations in palliative care research [35]. Despite this, there were frequent references to metaphysical concepts, such as the call for 'someone out there', suggesting that existential concerns may simply be taking new forms with the decline of traditional religious practices in UK society.

Also absent in Kate's Twitter narrative are 'battle metaphors', such as talk of fighting or being at war with cancer. While these are common in some cultural and clinical settings and have been the subject of research [36], experts have advised healthcare professionals against using them with patients, to avoid inducing feelings of failure for what is a biologically-determined outcome [37]. As a health professional, Kate Granger would have been acutely aware of her prognosis and may thus have chosen to focus on coping and preserving her quality of life. This also illustrates the need to recognise that individual patient characteristics can influence how a 'good death' is experienced [4].

Benefits of analysing social media data

This modest study takes a first step in demonstrating how these emerging data sources may elucidate terminal patients' physical and psychosocial responses during the illness trajectory and thus help to inform the provision of supportive and palliative care services at different stages. Analysing the social media postings of individuals like Kate Granger can provide a unique window into their 'lived experiences', including at highly emotional and sensitive stages, which can be difficult to access using conventional direct research methods. Social media are becoming the norm for communication amongst younger people and as these 'digital natives' progress to later stages of life, where death becomes more imminent, there are likely to be greater opportunities for such research.

Social media data are not only useful for focused studies, as we have undertaken here, but also present opportunities for research at scale. Automated tools for social media mining, natural language processing and sentiment analysis – such as those from Crimson Hexagon, Hootsuite, Symplur, Keyhole and Sproutsocial - are now widely used in the marketing sector and are transferable to academic research [38]. Other forms of digital data are already being used in this way. For example, in the only previous study we found to have charted the cancer trajectory for sarcoma, Tang and colleagues converged electronic records with individually-administered questionnaires, in order to profile levels of distress in as many as 74 patients before, during and after surgery, with the aim of understanding the psychological and socioeconomic factors influencing resilience, coping and outcomes [5]. A study currently underway at Stanford University [39] is using historical patient data to train Deep Learning algorithms to identify dying patients from their electronic health records and proactively bring them to the attention of palliative care staff in a hospital setting. Given such developments, it is not unreasonable to envisage a future in which data from patients' social media feeds, healthcare records and wearable monitoring devices are linked and processed using artificial intelligence, to generate real-time adaptive decision support for 'precision' palliative care.

Implications for supportive and palliative care

Illness trajectories have proven valuable as a means of describing the physical and psychosocial progression of cancer and other conditions [3]. While mapping the objective physical and medical aspects of these journeys is relatively straightforward, mapping their psychological and socio-emotional aspects typically requires in-depth qualitative studies with patients and their loved ones, which limits the usefulness of the findings for practitioners. Being able to study the journey towards death in the digital world opens a new window into the concerns, needs and vulnerabilities dying patients experience at different points in time, which may help to target the provision of supportive and palliative care, as well as enabling health professionals to understand patients' perspectives on the care they deliver.

The longitudinal data posted by Kate Granger provide evidence of the involvement of palliative care services during the patient journey. In Kate's case, she received palliative care relatively early, with her first tweet about this being posted on 25 October 2012, 14 months after her initial diagnosis. This is later supported by her tweet *"…had palliative care team support almost from day one. They've been amazing... See the whole of me"* – 4 May 2016. While palliative care services still tend to focus on the shorter dying phase [34] research has demonstrated that engagement over longer trajectories can enable better advanced planning for a good death, empower patients attempting to gain control over their illness and help to alleviate concerns about the possible nature of death [3]. All of these were evident in Kate's tweet *"After seeing my lovely palliative care nurse this a.m we've decided hospice admission for symptom control best course of action…"* - 8 July 2016.

End-of-life care planning must be multi-dimensional, with palliative care services playing the role of a mediator in helping patients to cope with their illness, optimising quality of life and achieving a dignified and peaceful death. Previous studies, including those focused on sarcoma patients, have revealed a range of problem-focused coping strategies, such as information seeking, choosing one's treatment team, and advocacy for oneself, as well as emotion-focused strategies such as support seeking, present-moment focus, distraction, denial and oversleeping [6]. Future analyses of patients' social media activity may help to verify or shed further light on these strategies, in addition to profiling illness trajectories and dimensions of the 'good death' framework. It may also help practitioners to better understand differences in patients' responses to their illness. For example, based on our analysis of Kate Granger's data, it appears that she did not use Twitter as a problem-oriented coping strategy, which might be explained by the focus on the last 6-months of her life rather than the period of initial diagnosis, or by her professional role as a geriatrician, which put her in a more informed role than most other patients. Understanding patients' trajectories of need can also help palliative care professionals to better anticipate and proactively mitigate distress [35]. In Kate's case this need for responsive approaches was evident in tweets such as *"Amazing care from #NHS today with my port flush, blood tests & psychology appointment…"* - 26 February 2016 and *"Psychology appt could not have come at a better time"* – 18 March 2016.

Another way in which digital ethnographic research methods may complement existing illness trajectory research, is by revealing the additional support provided to patients by their online social networks and communities. Kate Granger had many online followers, who helped to lift her spirits during periods of difficult treatment and distress, and studying the patterns of support and reciprocity in these digital spaces may suggest new ways in which to help patients nearing death. Despite Kate's predicament, she did not appear to be lonely, as can often be the case, and regularly thanked her followers for their support during acute episodes of crisis. As such, palliative care teams may consider recommending that terminal patients establish an online presence on social media and share their experiences with others, both as a therapeutic coping strategy and as a means of obtaining additional social support beyond their immediate family, friends and care services. Of course, it should be recognised that not all patients are able or willing to share their experiences online and many prefer to remain anonymous when doing so.

Social media also offer opportunities to study patients' *after* death. Services are emerging which enable users to preserve their digital legacy [40] through, for example, a social media auto-biography (e.g. deadsocial.org) or an avatar that draws on their social media history to engage in realistic interactions with friends and relatives (lifenaught.com) [41]. Although such tools may offer therapeutic value for patients preparing for a good death, or their bereaved loved ones, evidence of them doing so remains absent. This also raises new ethical and sociological questions about the responsible management of personas after death and the donation of social media archives for future research, in a similar way to the donation of medical records, tissue samples or body parts [42].

Limitations

The cohort size of one limits the extent to which the findings of this study can be generalised to other patient groups. We recommend further comparable research to extend the evidence-base, including studies exploring whether the end-of-life trajectories reported for different conditions in previous research are also evident on Twitter or other social media platforms [35].

During our study, we did not extract and analyse the entire sample of 12,500 tweets from the 4 years that the @GrangerKate Twitter account was active; limiting the conclusions drawn with regards to the complete illness trajectory from diagnosis to death. Aside from further manual analysis, this could be addressed using automated social media mining techniques and natural language processing tools. However, these approaches also have challenges, including variability in the types of social media to which they are suited, the scope and quality of the data available for analysis, and the cost of software and data access licences.

Despite numerous studies into what constitutes a good death, there is little agreement about its definition. Key features have been identified as: preferences for a specific dying process, pain-free status, religiosity/spirituality, emotional well-being, life completion, treatment preferences, dignity, family, quality of life and relationships with healthcare providers [43]. Although valuable and well respected, the 'framework for a good death' goes only some way towards accounting for the complexity of the end-of-life experience. In this study, the different dimensions were linked and therefore in some cases were difficult to distinguish as part of the analysis.

As already noted, the ethics of using potentially sensitive personal data placed on social media by individuals experiencing illness also represents a grey area, given the public nature of platforms like Twitter. Nevertheless, researchers should adhere with appropriate ethics guidelines, seek consent where appropriate and possible, and manage extracted data in a responsible and respectful way [44].

Conclusion

We believe this is the first study to have systematically analysed the end-of-life illness trajectory expressed in a patient's social media activity. Our results indicate that the data posted by terminal patients on Twitter can provide insights that may be comparable to, or compliment, those garnered using more traditional qualitative research techniques. To quote one of Kate's tweets, *"the power of patient narrative cannot be underestimated"* – 13 April 2016. While our analysis was at the structured end of the digital ethnographic spectrum, it nevertheless shows the value of such methods for understanding how terminal disease is experienced by and affects individuals, how they cope, how support is sought and obtained and how patients feel about the ability of palliative care services to meet their needs at different stages. Further research is warranted to extend this analysis across the wider trajectory of life-limiting illness and to a variety of disease types, as well as to explore the use of data mining and pattern recognition techniques to study larger cohorts and different social media platforms. As part of a wider agenda for 'palliative social media' [19] we recommend efforts to engage health professionals in exploring how digital end-of-life trajectories may inform the provision of supportive and palliative care, to improve the quality of life and death for patients like Kate.

Abbreviations

AoIR: Association of Internet Researchers; BPS: British Psychological Society; ERCP: Endoscopic Retrograde Cholangiopancreatographies; ESRC: Economic and Social Research Council; NHS: National Health Service

Acknowledgements

The authors would like to thank Kate Granger for tweeting about her patient experiences and Chris Pointon, her widower, for agreeing to our study and the publication of our results.

Funding

CP is a collaborator on three RCUK-funded research programmes relevant to this article: The Farr Institute for Health Informatics Research, Scotland, MRC Grant number MR/K007017/1; The Administrative Data Research Centre for Scotland, ESRC Grant number ES/L007487/1; The Science and Practice of Social Machines EPSRC Grant number EP/ J017728/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The data used in this study are publicly available via the Twitter account @GrangerKate.

Authors' contributions

JT and CP conceived and planned the study. JT extracted, screened and analysed the data, with input from CP. JT first-drafted the manuscript and CP edited and contributed to the writing. The authors received no specific funding for this work. Both authors read and approved the final manuscript.

Ethics approval and consent to participate

Ethics approval was provided by the University of Edinburgh's Institutional Review Board. Kate Granger was deceased at the time of the study but written agreement to extract, analyse and publish her tweets was sought and obtained from her widower Chris Pointon.

Consent for publication

The manuscript was shared with Chris Pointon and received his written approval.

Competing interests

This study was conducted as part of JT's self-funded PhD research project, supervised by CP. JT is also an employee of Ernst and Young Ltd. but the organization was not involved in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Publisher's Note

Author details

[1]eHealth Research Group, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [2]Ernst and Young AG, Basel, Switzerland.

Received: 24 April 2017 Accepted: 16 January 2018

References

1. Glaser BG and Strauss AL. Time for dying. Chicago: Aldine Pub. Co.; 1968.
2. Strauss LA, Glaser BC. Chronic illness and the quality of life. St. Louis: C.V. Mosby; 1975.
3. Murray SA, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. BMJ. 2005;330(7498):1007–11. 101136/bmj33074981007 2005
4. Emanuel EJ, Emanuel LL. The promise of a good death. Lancet (London, England). 1998;351(Suppl 2):SII21.
5. Tang MH, Castle DJ, Choong PFM. Identifying the prevalence, trajectory, and determinants of psychological distress in extremity sarcoma. Sarcoma. 2015; 2015:745163. Epub 2015 Feb 12. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4342175/.
6. Segal K, DuHamel KN, Posh L. Psychological Adaptation, Coping, and Distress in Adult-Onset Soft Tissue Sarcomas. ESUN A Periodical for the Sarcoma Community. 2010. http://sarcomahelp.org/coping.html.
7. Kaplan A, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Business Horizons. 2010;53(1):59–68.
8. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of health 2.0 and medicine 2.0: a systematic review. J Med Internet Res. 2010;12(2):e18.
9. Hilyard K, Broniatowski D, Dredzde M. How Far Can Twitter Reach in Good Survey Research? Social Science Space. 2015. https://www.socialsciencespace.com/2015/04/how-far-can-twitter-reach-in-good-survey-research/. Accessed 19 Jan 2018.
10. Pedersen ER, Kurz J. Using Facebook for health-related research study recruitment and program delivery. Current Opinion in Psychology. 2016;9:38.
11. Zooniverse. https://www.zooniverse.org.
12. Wilson RE, Gosling SD, Graham LT. A review of Facebook research in the social sciences. Perspect Psychol Sci. 2012;7(3):203–20.
13. Anstead N, O'Loughlin B. Social media analysis and public opinion: the 2010 UK general election. J Comput-Mediat Commun. 2015;20(2):204–20.
14. Murphy J, et al. Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research. American Association for Public Opinion Research. 2014. https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf. Accessed 19 Jan 2018.
15. Eysenbach G. Infodemiology and Infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res. 2009;11(1):e11.
16. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, Hartling L. Social media use among patients and caregivers: a scoping review. BMJ Open. 2013;3(5).
17. Gualtieri L, Akhtar FY. Cancer patient blogs: how patients, clinicians, and researchers learn from rich narratives of illness. In: Information technology interfaces (ITI), proceedings of the ITI 2013 35th international conference on: 24-27 June 2013; 2013. p. 3–8.
18. Tsuya A, Sugawara Y, Tanaka A, Narimatsu H. Do cancer patients tweet? Examining the twitter use of cancer patients in Japan. J Med Internet Res. 2014;16(5):e137.
19. Taubert M, Watts G, Boland J, Radbruch L. Palliative social media. BMJ Support Palliat Care. 2014;4(1):13.
20. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res. 2013;15(7):e147.
21. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. et al. Using social Media for Actionable Disease Surveillance and Outbreak Management: a systematic literature review. PLoS ONE 10(10): e0139701. doi:https://doi.org/10.1371/journal.pone.0139701/.
22. Kate Granger [https://en.wikipedia.org/wiki/Kate_Granger]. Accessed 19 Jan 2018.
23. @GrangerKate [https://twitter.com/grangerkate?lang=en]. Accessed 19 Jan 2018.
24. UK Department of Work and Pensions IB204. 2004. https://www.gpnotebook.co.uk/simplepage.cfm?ID=x20050803153747160230.
25. Varis PK. Digital Ethnography. Tilburg Papers in Culture Studies Paper 104 (online) Tilburg University. 2014. https://www.tilburguniversity.edu/upload/c428e18c-935f-4d12-8afb-652e19899a30_TPCS_104_Varis.pdf.
26. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277.
27. Hymes DH. Ethnography, linguistics, narrative inequality : toward an understanding of voice. London: Taylor & Francis; 1996.
28. Hookway N. Entering the blogosphere': some strategies for using blogs in social research. Qual Res. 2008;8(1):91–113.
29. Economic and Social Research Council (UK) Framework for Research Ethics. 2015. http://www.esrc.ac.uk/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2015/.
30. British Psychological Society. Report from the Working Party on Conducting Research on the Internet. Guidelines for ethical practice in conducting psychological research online. http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf.
31. Markham A, Buchanan E. Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee. Association of Internet Researchers. 2012. http://www.aoir.org/reports/ethics2.pdf.
32. Taylor J, Pagliari C. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? Research Ethics 2017: 1747016117738559. https://doi.org/10.1177/1747016117738559.
33. Barclay S, Froggatt K, Crang C, Mathie E, Handley M, Iliffe S, Manthorpe J, Gage H, Goodman C. Living in uncertain times: trajectories to death in residential care homes. Br J Gen Pract. 2014;64(626):e576.
34. Reed E, Corner J. Defining the illness trajectory of metastatic breast cancer. BMJ Support Palliat Care. 2015;5(4):358.
35. Murray SA, Kendall M, Grant E, Boyd K, Barclay S, Sheikh A. Patterns of social, psychological, and spiritual decline toward the end of life in lung cancer and heart failure. J Pain Symptom Manag. 2007;34(4):393–402.
36. Southall D. The patient's use of metaphor within a palliative care setting: theory, function and efficacy. A narrative literature review. Palliat Med. 2012; 27(4):304–13.
37. Gajewski M. May I take your metaphor? - how we talk about cancer. Cancer Research UK Science Blogs. 2015. http://scienceblog.cancerresearchuk.org/2015/09/28/may-i-take-your-metaphor-how-we-talk-about-cancer/.
38. Mindruta R. Top 15 Free Social Media Monitoring Tools. Brandwatch, Blog. 2017. https://www.brandwatch.com/blog/top-10-free-social-media-monitoring-tools/.
39. Avati A, Jung K, Harman S, Downing L, Ng A, Shah NH. Improving Palliative Care with deep learning IEEE international conference on bioinformatics and biomedicine 2017. [https://arxiv.org/pdf/1711.06402.pdf].
40. Cronin G. Living and dying on social media. CILIP Update. 2015. p. 35–37. http://www.cilip.org.uk/blog/living-dying-social-media.
41. Calvard J. Death and Grief Online - The Opportunities and Challenges of Incorporating Digital Legacies into Palliative Care in Hospice Settings. St. Columba's Hospice. 2016. http://www.stcolumbashospice.org.uk/wp-content/uploads/2016/11/Digital-legacy-presentation.pdf.
42. Shaw DM, Gross JV, Erren TC. Why You Should Donate Your Medical Data When You Die - Organs are not the only item of value from the deceased. Scientific American, The Conversation. 2017. https://www.scientificamerican.com/article/why-you-should-donate-your-medical-data-when-you-die/.
43. Meier EA, Gallegos JV, Thomas LP, Depp CA, Irwin SA, Jeste DV. Defining a good death (successful dying): literature review and a call for research and public dialogue. Am J Geriatr Psychiatry. 2016;24(4):261–71.
44. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res. 2013;15(4).

A.2.4.1 Additional information and material

Media coverage:

BMC Blog Series – Wake up, and smell the covfefe, Mark Taubert, 6 February 2018

https://blogs.biomedcentral.com/bmcseriesblog/2018/02/06/wake-smell-covfefe/

A.3 Results of formative search described in Chapter 2.2

Total number of publications by search query

Search query	**Total number of publications from 1997-2017**
"social media" AND (health OR illness OR disease) AND human	3425
digital AND surveillance AND (health OR illness OR disease) AND human	2935
facebook AND (health OR illness OR disease) AND human	939
twitter AND (health OR illness OR disease) AND human	709
"social media" AND surveillance AND (health OR illness OR disease) AND human	665
blog AND (health OR illness OR disease) AND human	602
"social media" AND epidemiology AND (disease OR health OR illness) AND human	581
"social network analysis" AND (disease OR health OR illness) AND human	462
youtube AND (health OR illness OR disease) AND human	351
"web 2.0" AND (health OR illness OR disease) AND human	312
facebook AND surveillance AND (health OR illness OR disease) AND human	206
twitter AND surveillance AND (health OR illness OR disease) AND human	187
digital AND ethnography AND (health OR illness OR disease) AND human	110
"online forum" AND (health OR illness OR disease) AND human	103
blog AND surveillance AND (health OR illness OR disease) AND human	76
patientslikeme AND (health OR illness OR disease) AND human	60
"online monitoring" AND (disease OR health OR illness) AND human	48
"health 2.0" AND (health OR illness OR disease) AND human	46
twitter AND mining AND (health OR illness OR disease) AND human	43
instagram AND (health OR illness OR disease) AND human	40
online AND "social network analysis" AND (disease OR health OR illness) AND human	29
"online forum" AND surveillance AND (health OR illness OR disease) AND human	25
"online community" AND surveillance AND (health OR illness OR disease) AND human	25
infoveillance AND (disease OR health OR illness) AND human	24
"online surveillance" AND (disease OR health OR illness) AND human	23
"social network analysis" AND web AND (disease OR health OR illness) AND human	23
"digital surveillance" AND (disease OR health OR illness) AND human	16
netnography AND (disease OR health OR illness) AND human	14
"medicine 2.0" AND (health OR illness OR disease) AND human	14
"social media analysis" AND (disease OR health OR illness) AND human	12
"online tracking" AND (disease OR health OR illness) AND human	12
"web analytics" AND (disease OR health OR illness) AND human	10
"social media analytics" AND (disease OR health OR illness) AND human	9
"social media monitoring" AND (disease OR health OR illness) AND human	9
blog AND mining AND (health OR illness OR disease) AND human	9
"web mining" AND (disease OR health OR illness) AND human	8
"digital monitoring" AND (disease OR health OR illness) AND human	7

Search query	**Total number of publications from 1997-2017**
facebook AND mining AND (health OR illness OR disease) AND human	7
"social media mining" AND (disease OR health OR illness) AND human	6
"web monitoring" AND (disease OR health OR illness) AND human	6
"social media platform" AND surveillance AND (health OR illness OR disease) AND human	6
"social media surveillance" AND (disease OR health OR illness) AND human	4
"web tracking" AND (disease OR health OR illness) AND human	4
digital AND "social network analysis" AND (disease OR health OR illness) AND human	4
"social network analysis" AND digital AND (disease OR health OR illness) AND human	4
"digital ethnography" AND (health OR illness OR disease) AND human	4
"digital tracking" AND (disease OR health OR illness) AND human	3
"online analytics" AND (disease OR health OR illness) AND human	3
"social media site" AND surveillance AND (health OR illness OR disease) AND human	3
"social media listening" AND (disease OR health OR illness) AND human	2
"social network analytics" AND (disease OR health OR illness) AND human	2
"social media website" AND surveillance AND (health OR illness OR disease) AND human	2
"web surveillance" AND (disease OR health OR illness) AND human	0
"social network surveillance" AND (disease OR health OR illness) AND human	0
"social media observing" AND (disease OR health OR illness) AND human	0
"social media tracking" AND (disease OR health OR illness) AND human	0
"digital observing" AND (disease OR health OR illness) AND human	0
"digital analytics" AND (disease OR health OR illness) AND human	0
"digital mining" AND (disease OR health OR illness) AND human	0
"digital listening" AND (disease OR health OR illness) AND human	0
"web observing" AND (disease OR health OR illness) AND human	0
"web listening" AND (disease OR health OR illness) AND human	0
"online observing" AND (disease OR health OR illness) AND human	0
"online mining" AND (disease OR health OR illness) AND human	0
"online listening" AND (disease OR health OR illness) AND human	0
"social network observing" AND (disease OR health OR illness) AND human	0
"social network mining" AND (disease OR health OR illness) AND human	0
"social network monitoring" AND (disease OR health OR illness) AND human	0
"social network listening" AND (disease OR health OR illness) AND human	0
"social network tracking" AND (disease OR health OR illness) AND human	0
linkedin AND (health OR illness OR disease) AND human	0
"sixdegrees" AND (health OR illness OR disease) AND human	0

Bibliography

1. Oh, H., et al., *What is eHealth (3): a systematic review of published definitions.* Journal of Medical Internet Research, 2005. **7**(1): p. e1.
2. Pagliari, C., et al., *What is eHealth (4): a scoping exercise to map the field.* Journal of Medical Internet Research, 2005. **7**(1): p. e9.
3. WHA. *Digital Health*. Seventy-First World Health Assembly Agenda Item 12.4 2018 [cited 2020; Archived at: http://archive.is/odo7u]. Available from: https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_R7-en.pdf.
4. WHO. *Draft Global Strategy on Digital Health 2020-2024*. 2020 [cited 2020; Archived at: http://archive.is/Jo7SO]. Available from: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf?sfvrsn=f112ede5_42.
5. **Kaplan, A.a.H., M**, *Users of the world, unite! The challenges and opportunities of Social Media.* Business Horizons, 2010. **53**(1): p. 59-68.
6. *LinkedIn*. [cited 2020; Archived at: http://archive.is/3O8sZ]. Available from: https://www.linkedin.com/.
7. *Facebook*. [cited 2020; Archived at: http://archive.is/XRTtb]. Available from: https://www.facebook.com.
8. *Twitter*. Archived at: http://www.webcitation.org/7534cfLqs]. Available from: http://www.twitter.com.
9. *YouTube*. [cited 2020; Archived at: http://archive.is/TPJd2]. Available from: https://www.youtube.com.
10. PatientsLikeMe. 2014; Archived at: http://www.webcitation.org/6Sa5rVOBU]. Available from: http://www.patientslikeme.com.
11. TuDiabetes. 2014; Archived at: http://www.webcitation.org/6Sa5bnrC4]. Available from: http://www.tudiabetes.org/profiles/members/.
12. *Support Communities*. [cited 2020; Archived at: http://archive.is/noBrs]. Available from: https://www.medhelp.org/forums/list.
13. Jacobs, D. *The use of Social Networks for E-Health Literature Synthesis*. Archived at: http://www.webcitation.org/6PXFl2nQ6]. Available from: http://people.cs.uct.ac.za/~djacobs/downloads/Literature_Review_JCBDAL001.pdf.
14. NIH. *National Library of Medicine - Medical Subject Headings, 2014 MeSH. MeSH Descriptor Data, Social Media*. 2012; Archived at: http://www.webcitation.org/6P052ld27]. Available from: https://www.nlm.nih.gov/cgi/mesh/2014/MB_cgi?mode=&index=26465&view=concept.
15. Lachman, V.D., *Social media: managing the ethical issues*. 2013, Jannetti Publications, Inc. p. 326.
16. Hilyard, K., D. Broniatowski, and M. Dredze. *How Far Can Twitter Reach in Good Survey Research?* 2015; Archived at: http://www.webcitation.org/6cXWhtMn3]. Available from: http://www.socialsciencespace.com/2015/04/how-far-can-twitter-reach-in-good-survey-research/.
17. Pedersen, E.R. and J. Kurz, *Using Facebook for Health-related Research Study Recruitment and Program Delivery.* Current Opinion in Psychology, 2016.
18. Trisha, G., *Citizen science: Amateur experts.* Nature, 2013. **496**(7444): p. 259.
19. Russell, D.J., et al., *Knowledge Exchange and Discovery in the Age of Social Media: The Journey From Inception to Establishment of a Parent-Led Web-Based Research Advisory Community for Childhood Disability.* Journal of Medical Internet Research, 2016. **18**(11): p. e293.
20. INVOLVE. *Guidance on the use of social media to actively involve people in research*. 2014; Archived at: http://www.webcitation.org/6cMnLbQK6]. Available from: http://www.invo.org.uk/wp-content/uploads/2014/11/9982-Social-Media-Guide-WEB.pdf.
21. Wilson, R.E., S.D. Gosling, and L.T. Graham, *A Review of Facebook Research in the Social Sciences.* Perspectives on Psychological Science, 2012. **7**(3): p. 203-220.
22. Anstead, N. and B. O'Loughlin, *Social Media Analysis and Public Opinion: The 2010 UK General Election.* Journal of Computer-Mediated Communication, 2015. **20**(2): p. 204-220.
23. Murphy, J., et al. *Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research*. 2014; Archived at: http://www.webcitation.org/6cXXitYw6]. Available from:

https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf.

24. Smith, C. *Social Networks Are Only Just Getting Started In Mining User Data*. 2014; Archived at:]. Available from: http://www.businessinsider.com/social-medias-big-data-future-2014-2?IR=T.
25. Bjerglund Andersen, N. and T. Söderqvist. *Social Media and Public Health Research*. 2012; Archived at: http://www.webcitation.org/6cCEV4nTm]. Available from: https://bjerglund.files.wordpress.com/2012/11/final-social-media-and-public-health-research1.pdf.
26. Eke, P.I., *Using Social Media for Research and Public Health Surveillance.* Journal of Dental Research, 2011. **90**(9): p. 1045-1046.
27. Eysenbach, G., *Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet.* J Med Internet Res, 2009. **11**(1): p. e11.
28. Choi, B.C.K., *The past, present, and future of public health surveillance.* Scientifica, 2012. **2012**: p. 875253-875253.
29. Ahmed, S.M., et al., *Clustering of chronic non-communicable disease risk factors among selected Asian populations: levels and determinants.* Global Health Action, 2009. **2**: p. 68-75.
30. Al-Nsour, M., et al., *Prevalence of Selected Chronic, Noncommunicable Disease Risk Factors in Jordan: Results of the 2007 Jordan Behavioral Risk Factor Surveillance Survey.* PREVENTING CHRONIC DISEASE, 2012. **9**.
31. Deepa, M., et al., *Noncommunicable Diseases Risk Factor Surveillance: Experience and Challenge from India.* Indian Journal of Community Medicine, 2011. **36**: p. S50-S56.
32. Ng, N., et al., *Using the INDEPTH HDSS to build capacity for chronic non-communicable disease risk factor surveillance in low and middle-income countries.* Global Health Action, 2009. **2**: p. 7-18.
33. Yichong, L., et al., *Risk factors for noncommunicable chronic diseases in women in China: surveillance efforts.* Los factores de riesgo de las enfermedades crónicas no transmisibles en las mujeres chinas: los esfuerzos de vigilancia., 2013. **91**(9): p. 650-660.
34. WHO. *Noncommunicable diseases*. 2013; Archived at: http://www.webcitation.org/6PqvwlZCv]. Available from: http://www.who.int/mediacentre/factsheets/fs355/en/.
35. WHO. *NCD Surveillance*. Archived at: http://www.webcitation.org/6PjUXY8VO]. Available from: http://www.who.int/ncd_surveillance/en/.
36. Velasco, E., et al., *Social Media and Internet-Based Data in Global Systems for Public Health Surveillance: A Systematic Review.* Milbank Quarterly, 2014. **92**(1): p. 7-33.
37. Eggleston, E.M., E.R. Weitzman, and *Innovative uses of electronic health records and social media for public health surveillance* . Current diabetes reports, 2014. **14**(3): p. 468-468.
38. Boman, M., et al. *Syndromic surveillance in the United Arab Emirates*. 2012.
39. Agheneza, T. *A systematic literature review on event-based public health surveillance systems*. 2011; Archived at: http://www.webcitation.org/6xpukEcdu]. Available from: http://edoc.sub.uni-hamburg.de/haw/volltexte/2012/1614/pdf/lsab12_27.pdf.
40. Fung, I.C.-H., et al., *Ebola virus disease and social media: A systematic review.* AJIC: American Journal of Infection Control, 2016. **44**(12): p. 1660-1671.
41. Guy, S., et al., *Social media: A systematic review to understand the evidence and application in infodemiology*. 2012. p. 1-8.
42. Bernardo, T.M., et al., *Scoping review on search queries and social media for disease surveillance: A chronology of innovation.* Journal of Medical Internet Research, 2013. **15**(7).
43. Velasco, E., et al., *Social Media and Internet-Based Data in Global Systems for Public Health Surveillance: A Systematic Review.* The Milbank Quarterly, 2014. **92**(1): p. 7-33.
44. Charles-Smith, L.E., et al., *Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review.* PLoS One, 2015. **10**(10): p. e0139701.
45. Al-Garadi, M.A., et al., *Using online social networks to track a pandemic: A systematic review.* J Biomed Inform, 2016. **62**: p. 1-11.
46. Zhou, X., et al., *Surveillance and response systems for elimination of tropical diseases: summary of a thematic series in Infectious Diseases of Poverty.* Infect Dis Poverty, 2016. **5**(1): p. 49.
47. O'Shea, J., *Digital disease detection: A systematic review of event-based internet biosurveillance systems.* International Journal of Medical Informatics, 2017. **101**: p. 15-22.

48. Yan, S.J., A.A. Chughtai, and C.R. Macintyre, *Utility and potential of rapid epidemic intelligence from internet-based sources.* Int J Infect Dis, 2017. **63**: p. 77-87.
49. Jones, E., et al., *Social networking and understanding alcohol-associated risk for people with type 1 diabetes: Friend or foe?* Diabetes Technology and Therapeutics, 2013. **15**(4): p. 308-314.
50. Cotter, A.P., et al., *Internet interventions to support lifestyle modification for diabetes management: A systematic review of the evidence.* Journal of Diabetes and its Complications, 2014. **28**(2): p. 243-251.
51. Nour, M., et al., *A Narrative Review of Social Media and Game-Based Nutrition Interventions Targeted at Young Adults.* Journal of the Academy of Nutrition and Dietetics, 2017. **117**(5): p. 735-752.e10.
52. Rose, T., et al., *A Systematic Review of Digital Interventions for Improving the Diet and Physical Activity Behaviors of Adolescents.* J Adolesc Health, 2017. **61**(6): p. 669-677.
53. Park, B.K.M.S.R.N. and C.P.R.N. Calamaro, *A Systematic Review of Social Networking Sites: Innovative Platforms for Health Research Targeting Adolescents and Young Adults.* Journal of Nursing Scholarship, 2013. **45**(3): p. 256-64.
54. Best, P., R. Manktelow, and B. Taylor, *Online communication, social media and adolescent wellbeing: A systematic narrative review.* Children and Youth Services Review, 2014. **41**: p. 27-36.
55. Baker, D.A. and G.P. Algorta, *The Relationship between Online Social Networking and Depression: A Systematic Review of Quantitative Studies.* Cyberpsychology, Behavior, and Social Networking, 2016. **19**(11): p. 638-648.
56. Robinson, J., et al., *Social media and suicide prevention: A systematic review.* Early Intervention in Psychiatry, 2016. **10**(2): p. 103-121.
57. Välimäki, M., et al., *Effectiveness of social media interventions for people with schizophrenia: A systematic review and meta-analysis.* Journal of Medical Internet Research, 2016. **18**(4).
58. Calvo, R.A., et al., *Natural language processing in mental health applications using non-clinical texts.* Natural Language Engineering, 2017. **23**(5): p. 649-685.
59. Marchant, A., et al., *A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown.* PloS one, 2017. **12**(8): p. e0181722.
60. Twomey, C. and G. O'Reilly, *Associations of Self-Presentation on Facebook with Mental Health and Personality Variables: A Systematic Review.* Cyberpsychology, Behavior, and Social Networking, 2017. **20**(10): p. 587-595.
61. Wongkoblap, A., M.A. Vadillo, and V. Curcin, *Researching Mental Health Disorders in the Era of Social Media: Systematic Review.* J Med Internet Res, 2017. **19**(6): p. e228.
62. Hamm, M.P., et al., *A systematic review of the use and effectiveness of social media in child health.* BMC Pediatrics, 2014. **14**(1).
63. Odone, A., et al., *Effectiveness of interventions that apply new media to improve vaccine uptake and vaccine coverage: A systematic review.* Human Vaccines and Immunotherapeutics, 2015. **11**(1): p. 72-82.
64. Kazemi, D.M., et al., *Systematic review of surveillance by social media platforms for illicit drug use.* J Public Health (Oxf), 2017: p. 1-14.
65. Golder, S., G. Norman, and Y.K. Loke, *Systematic review on the prevalence, frequency and comparative value of adverse events data in social media.* Br J Clin Pharmacol, 2015. **80**(4): p. 878-88.
66. Emmert, M., U. Sander, and F. Pisch, *Eight Questions About Physician-Rating Websites: A Systematic Review.* J Med Internet Res, 2013. **15**(2): p. e24.
67. Smailhodzic, E., et al., *Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals.* BMC Health Services Research, 2016. **16**(1).
68. Verhoef, L.M., et al., *Social media and rating sites as tools to understanding quality of care: a scoping review.* J Med Internet Res, 2014. **16**(2): p. e56.
69. Hamm, M.P., et al., *Social media use among patients and caregivers: a scoping review.* BMJ OPEN, 2013. **3**(5).
70. Tursunbayeva, A., M. Franco, and C. Pagliari, *Use of social media for e-Government in the public health sector: A systematic review of published studies.* Government Information Quarterly, 2017. **34**(2): p. 270-282.
71. Smith, T. and R. Lambert, *A systematic review investigating the use of twitter and facebook in university-based healthcare education.* Health Education, 2014. **114**(5): p. 347-366.

72. Golder, S., et al., *Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review.* J Med Internet Res, 2017. **19**(6): p. e195.
73. Conway, M., *Ethical issues in using twitter for public health surveillance and research: Developing a taxonomy of ethical concepts from the research literature.* Journal of Medical Internet Research, 2014. **16**(12).
74. Gustafson, D.L. and C.F. Woodworth, *Methodological and ethical issues in research using social media: A metamethod of Human Papillomavirus vaccine studies.* BMC Medical Research Methodology, 2014. **14**(1).
75. Moorhead, S.A., et al., *A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication.* J Med Internet Res, 2013. **15**(4): p. e85.
76. Capurro, D., et al., *The use of social networking sites for public health practice and research: a systematic review.* J Med Internet Res, 2014. **16**(3): p. e79.
77. Laaksonen, C., H. Jalonen, and J. Paavola, *Utilising Social Media for Intervening and Predicting Future Health in Societies*. 2014, Springer Verlag. p. 100-108.
78. Hu, Y., *Health communication research in the digital age: A systematic review.* Journal of Communication in Healthcare, 2015. **8**(4): p. 260-288.
79. Gabarron, E. and R. Wynn, *Use of social media for sexual health promotion: A scoping review.* Global Health Action, 2016. **9**(1).
80. Carson, K.V., et al., *Mass media interventions for preventing smoking in young people.* Cochrane Database Syst Rev, 2017. **6**: p. Cd001006.
81. Daher, J., et al., *Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017).* BMJ Open, 2017. **7**(11): p. e017604.
82. Overbey, K.N., L.A. Jaykus, and B.J. Chapman, *A systematic review of the use of social media for food safety risk communication.* Journal of Food Protection, 2017. **80**(9): p. 1537-1549.
83. Sinnenberg, L., et al., *Twitter as a Tool for Health Research: A Systematic Review.* American journal of public health, 2017. **107**(1): p. 143.
84. Whitaker, C., S. Stevelink, and N. Fear, *The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review.* J Med Internet Res, 2017. **19**(8): p. e290.
85. Carvalho, R.E.V. and C.T. Lopes. *Social network analysis to understand behaviour dynamics in online health communities: A systematic review*. 2016. IEEE Computer Society.
86. Patel, R., et al., *Social Media Use in Chronic Disease: A Systematic Review and Novel Taxonomy.* Am J Med., 2015. **128**(12): p. 1335-1350.
87. Fung, I.C., et al., *Ebola virus disease and social media: A systematic review*. (1527-3296 (Electronic)).
88. Coto, M., et al., *Social media and elderly people: Research trends*, G. Meiselwitz, Editor. 2017, Springer Verlag. p. 65-81.
89. Hamm, M.P., et al., *Social Media Use by Health Care Professionals and Trainees: A Scoping Review.* ACADEMIC MEDICINE, 2013. **88**(9): p. 1376-1383.
90. Arksey, H. and L. O'Malley, *Scoping studies: towards a methodological framework.* International Journal of Social Research Methodology, 2005. **8**(1): p. 19-32.
91. Levac, D., H. Colquhoun, and K.K. O'Brien, *Scoping studies: advancing the methodology.* Implementation Science : IS, 2010. **5**: p. 69-69.
92. Peters MDJ, G.C., McInerney P, Baldini Soares C, Khalil H, Parker D. . *Chapter 11: Scoping Reviews.* . Joanna Briggs Institute Reviewer's Manual. 2017; Available from: Available from https://reviewersmanual.joannabriggs.org/
93. Armstrong, R., et al., *Cochrane Update. 'Scoping the scope' of a cochrane review.* (1741-3850 (Electronic)).
94. Lam, M.T. and M. McDiarmid, *Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014.* Journal of the Medical Library Association : JMLA, 2016. **104**(4): p. 284-289.
95. Moher, D., et al., *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.* PLoS Clinical Trials, 2009. **6**(7): p. 1-6.
96. SJR. *Scimago Journal & Country Rank*. 2016; Archived at: http://www.webcitation.org/6xpu101nf]. Available from: http://www.scimagojr.com/journalrank.php.
97. *World Bank Country and Lending Groups*. Data 2017; Archived at: http://www.webcitation.org/6q87seiSy]. Available from:

https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.
98. Injadat, M., F. Salo, and A.B. Nassif, *Data mining techniques in social media: A survey.* Neurocomputing, 2016. **214**(C): p. 654-670.
99. Conway, M., *Ethical Issues in Using Twitter for Public Health Surveillance and Research: Developing a Taxonomy of Ethical Concepts From the Research Literature.* J Med Internet Res, 2014. **16**(12): p. e290.
100. Abedin, T., et al., *Social Media as a Platform for Information About Diabetes Foot Care: A Study of Facebook Groups.* Canadian Journal of Diabetes, 2017. **41**(1): p. 97-101.
101. Abramson, K., B. Keefe, and W.Y.S. Chou, *Communicating about cancer through Facebook: A qualitative analysis of a breast cancer awareness page.* Journal of Health Communication, 2015. **20**(2): p. 237-243.
102. Akay, A., A. Dragomir, and B. Erlandsson. *A novel data-mining approach leveraging social media to monitor and respond to outcomes of diabetes drugs and treatment*. in *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*. 2013.
103. Akay, A., A. Dragomir, and B. Erlandsson. *A novel data-mining platform leveraging social media to monitor outcomes of Januvia*. in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013.
104. Akay, A., A. Dragomir, and B.E. Erlandsson, *Network-based modeling and intelligent data mining of social media for improving care.* IEEE Journal of Biomedical and Health Informatics, 2015. **19**(1): p. 210-218.
105. Akay, A., A. Dragomir, and B.E. Erlandsson, *A novel data-mining approach leveraging social media to monitor consumer opinion of Sitagliptin.* IEEE Journal of Biomedical and Health Informatics, 2015. **19**(1): p. 389-396.
106. AlQarni, Z.A., F. Yunus, and M.S. Househ, *Health information sharing on Facebook: An exploratory study on diabetes mellitus.* Journal of Infection and Public Health, 2016. **9**(6): p. 708-712.
107. Azer, S.A., et al., *Accuracy and readability of cardiovascular entries on Wikipedia: Are they reliable learning resources for medical students?* BMJ Open, 2015. **5**(10).
108. Beusterien, K., et al., *Real-world experience with colorectal cancer chemotherapies: Patient web forum analysis.* ecancermedicalscience, 2013. **7**(1).
109. Bian, J., et al., *Using social media data to understand the impact of promotional information on laypeople's discussions:a case study of lynch syndrome.* Journal of Medical Internet Research, 2017. **19**(12).
110. Biyani, P., et al. *Co-training over Domain-independent and Domain-dependent features for sentiment analysis of an online cancer support community*. in *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. 2013.
111. Blank, T.O. and M. Adams-Blodnieks, *The who and the what of usage of two cancer online communities.* Computers in Human Behavior, 2007. **23**(3): p. 1249-1257.
112. Bond, C.S. and O.H. Ahmed, *Can I help you? Information sharing in online discussion forums by people living with a long-term condition.* Journal of innovation in health informatics, 2016. **23**(3): p. 620-626.
113. Borgmann, H., et al., *Activity, content, contributors, and influencers of the twitter discussion on urologic oncology.* Urologic Oncology: Seminars and Original Investigations, 2016. **34**(9): p. 377-383.
114. Bosley, J.C., et al., *Decoding twitter: Surveillance and trends for cardiac arrest and resuscitation communication.* Resuscitation, 2013. **84**(2): p. 206-212.
115. Bravo, C.A. and L. Hoffman-Goetz, *Tweeting About Prostate and Testicular Cancers: What Are Individuals Saying in Their Discussions About the 2013 Movember Canada Campaign?* Journal of Cancer Education, 2016. **31**(3): p. 559-566.
116. Bravo, C.A. and L. Hoffman-Goetz, *Social Media and Men's Health: A Content Analysis of Twitter Conversations During the 2013 Movember Campaigns in the United States, Canada, and the United Kingdom.* American Journal of Men's Health, 2017. **11**(6): p. 1627-1641.
117. Bui, N., J. Yen, and V. Honavar, *Temporal causality analysis of sentiment change in a cancer survivor network.* IEEE Transactions on Computational Social Systems, 2016. **3**(2): p. 75-87.
118. Buis, L.R. and P. Whitten, *Comparison of social support content within online communities for high-and low-survival-rate cancers.* CIN - Computers Informatics Nursing, 2011. **29**(8): p. 461-467.

119. Campbell, K.A., N.S. Coulson, and H. Buchanan, *Empowering processes within prostate cancer online support groups.* International Journal of Web Based Communities, 2013. **9**(1): p. 51-66.
120. Chen, A.T., *Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups.* Patient Education and Counseling, 2012. **87**(2): p. 250-257.
121. Chomutare, T., E. Årsand, and G. Hartvigsen. *Mining Symptoms of Severe Mood Disorders in Large Internet Communities*. in *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*. 2015.
122. Chul, L.K., et al., *Transmissibility of the campaign for colorectal cancer awareness in Korea among twitter users.* Annals of Coloproctology, 2016. **32**(5): p. 184-189.
123. Chung, J.E., *Retweeting in health promotion: Analysis of tweets about Breast Cancer Awareness Month.* Computers in Human Behavior, 2017. **74**: p. 112-119.
124. Coll-Planas, G. and M. Visa, *The wounded blogger: analysis of narratives by women with breast cancer.* Sociology of Health and Illness, 2016. **38**(6): p. 884-898.
125. Coloma, P.M., et al., *Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies.* Drug Safety, 2015. **38**(10): p. 921-930.
126. Crannell, W.C., et al., *A pattern-matched Twitter analysis of US cancer-patient sentiments.* Journal of Surgical Research, 2016. **206**(2): p. 536-542.
127. Crook, B., et al., *Hanging by a thread: exploring the features of nonresponse in an online young adult cancer survivorship support community.* Journal of Cancer Survivorship, 2016. **10**(1): p. 185-193.
128. Curtis, J.R., et al., *Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-to-consumer advertising.* Arthritis Research and Therapy, 2017. **19**(1).
129. Dai, H., B.R. Lee, and J. Hao, *Predicting Asthma Prevalence by Linking Social Media Data and Traditional Surveys.* Annals of the American Academy of Political and Social Science, 2017. **669**(1): p. 75-92.
130. Derksen, C., et al., *"what say ye gout experts?" a content analysis of questions about gout posted on the social news website Reddit.* BMC Musculoskeletal Disorders, 2017. **18**(1).
131. Diddi, P. and L.K. Lundy, *Organizational Twitter Use: Content Analysis of Tweets during Breast Cancer Awareness Month.* Journal of Health Communication, 2017. **22**(3): p. 243-253.
132. Donelle, L. and L. Hoffman-Goetz, *Functional health literacy and cancer care conversations in online forums for retired persons.* Informatics for Health and Social Care, 2009. **34**(1): p. 59-72.
133. Eichstaedt, J.C., et al., *Psychological Language on Twitter Predicts County-Level Heart Disease Mortality.* Psychological Science, 2015. **26**(2): p. 159-169.
134. Elhadad, N., et al., *Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions.* AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium, 2014. **2014**: p. 516-525.
135. Erfani, S.S., B. Abedin, and F. Daneshgar. *A qualitative evaluation of communication in Ovarian Cancer Facebook communities*. 2013. IEEE Computer Society.
136. Freedman, R.A., et al., *Learning from social media: utilizing advanced data extraction techniques to understand barriers to breast cancer treatment.* Breast Cancer Research and Treatment, 2016. **158**(2): p. 395-405.
137. Frost, J., et al., *Patient-reported outcomes as a source of evidence in off-label prescribing: Analysis of data from PatientsLikeMe.* Journal of Medical Internet Research, 2011. **13**(1).
138. Gabarron, E., A. Makhlysheva, and L. Marco, *Type 1 Diabetes in Twitter: Who All Listen To?,* A. Georgiou, I.N. Sarkar, and P.M. de Azevedo Marques, Editors. 2015, IOS Press. p. 972.
139. Gage-Bouchard, E.A., et al., *Cancer Communication on Social Media: Examining How Cancer Caregivers Use Facebook for Cancer-Related Communication.* Cancer Nursing, 2017. **40**(4): p. 332-338.
140. Gonzalez-Polledo, E., *Chronic Media Worlds: Social Media and the Problem of Pain Communication on Tumblr.* Social Media and Society, 2016. **2**(1).
141. Gonzalez-Polledo, E. and J. Tarr, *The thing about pain: The remaking of illness narratives in chronic pain expressions on social media.* New Media and Society, 2016. **18**(8): p. 1455-1472.
142. Goranson, A., et al., *Dying Is Unexpectedly Positive.* Psychological Science, 2017. **28**(7): p. 988-999.

143. Greene, J.A., et al., *Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook.* Journal of General Internal Medicine, 2011. **26**(3): p. 287-292.
144. Greenwood, M., et al. *Automatic Extraction of Personal Experiences from Patients' Blogs: A Case Study in Chronic Obstructive Pulmonary Disease*. in *2013 International Conference on Cloud and Green Computing*. 2013.
145. Hadert, A. and K. Rodham, *The invisible reality of arthritis: A qualitative analysis of an online message board.* Musculoskeletal Care, 2008. **6**(3): p. 181-196.
146. Hale, T.M., et al., *Representation of health conditions on Facebook: Content analysis and evaluation of user engagement.* Journal of Medical Internet Research, 2014. **16**(8): p. e182.
147. Hamon, T. and R. Gagnayre, *Improving knowledge of patient skills thanks to automatic analysis of online discussions.* Patient Education and Counseling, 2013. **92**(2): p. 197-204.
148. Harris, J.K., et al., *Diabetes topics associated with engagement on twitter.* Preventing Chronic Disease, 2015. **12**(5).
149. Hewitt-Taylor, J. and C.S. Bond, *What E-patients want from the doctor-patient relationship: Content analysis of posts on discussion boards.* Journal of Medical Internet Research, 2012. **14**(6).
150. Hoedebecke, K., et al., *Health care and social media: What patients really understand.* F1000Research, 2017. **6**.
151. Holmberg, K., K. Eriksson-Backa, and S. Ek, *Tweeting about Diabetes and Diets - Content and Conversational Connections*. 2014, Springer Verlag. p. 46-56.
152. Huh, J., M. Yetisgen-Yildiz, and W. Pratt, *Text classification for assisting moderators in online health communities.* Journal of Biomedical Informatics, 2013. **46**(6): p. 998-1005.
153. Kanthawala, S., et al., *Answers to health questions: Internet search results versus online health community responses.* Journal of Medical Internet Research, 2016. **18**(4).
154. Keim-Malpass, J. and R.H. Steeves, *Talking With death at a diner: Young women's online narratives of cancer.* Oncology Nursing Forum, 2012. **39**(4): p. 373-378+406.
155. Keim-Malpass, J., et al., *Young Women's Experiences With Complementary Therapies During Cancer Described Through Illness Blogs.* Western Journal of Nursing Research, 2013. **35**(10): p. 1309-1324.
156. Keim-Malpass, J., et al., *Blogging through cancer: Young women's persistent problems shared online.* Cancer Nursing, 2013. **36**(2): p. 163-172.
157. Keim-Malpass, J., et al., *"It's Back! My Remission Is Over": Online Communication of Disease Progression Among Adolescents With Cancer.* Journal of Pediatric Oncology Nursing, 2016. **33**(3): p. 209-217.
158. Kent, E.E., et al., *"Obesity is the New Major Cause of Cancer": Connections Between Obesity and Cancer on Facebook and Twitter.* Journal of Cancer Education, 2016. **31**(3): p. 453-459.
159. Khasnavis, S., A.B. Rosenkrantz, and V. Prabhu, *Using Twitter to Assess the Public Response to the United States Preventive Services Task Force Guidelines on Lung Cancer Screening with Low Dose Chest CT.* Journal of Digital Imaging, 2017. **30**(3): p. 323-327.
160. Kim, B. and D.M. Gillham, *The experience of young adult cancer patients described through online narratives.* Cancer Nursing, 2013. **36**(5): p. 377-384.
161. Kim, B. and D. Gillham, *Gender differences among young adult cancer patients: A study of blogs.* CIN - Computers Informatics Nursing, 2015. **33**(1): p. 3-9.
162. Koschack, J., et al., *Scientific versus experiential evidence: Discourse analysis of the chronic cerebrospinal venous insufficiency debate in a multiple sclerosis forum.* Journal of Medical Internet Research, 2015. **17**(7).
163. Koteyko, N. and D. Hunt, *Performing health identities on social media: An online observation of Facebook profiles.* Discourse, Context and Media, 2016. **12**: p. 59-67.
164. Krittanawong, C., et al., *Tweeting influenza vaccine to cardiovascular health community.* European Journal of Cardiovascular Nursing, 2017. **16**(8): p. 704-706.
165. Lee, K., A. Agrawal, and A. Choudhary, *Real-time disease surveillance using Twitter data: demonstration on flu and cancer*, in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013, ACM: Chicago, Illinois, USA.
166. Lenoir, P., et al., *Raising Awareness About Cervical Cancer Using Twitter: Content Analysis of the 2015 #SmearForSmear Campaign.* Journal of medical Internet research, 2017. **19**(10): p. e344.
167. Liu, X. and H. Chen, *Identifying adverse drug events from patient social media: A case study for diabetes.* IEEE Intelligent Systems, 2015. **30**(3): p. 44-51.

168. Lobchuk, M., et al., *A Qualitative Analysis of "naturalistic" Conversations in a Peer-Led Online Support Community for Lung Cancer.* Cancer Nursing, 2015. **38**(5): p. E21-E31.
169. Loeb, S., et al., *Tweet this: how advocacy for breast and prostate cancers stacks up on social media.* BJU Int, 2017. **120**(4): p. 461-463.
170. Love, B., et al., *Work and "mass personal" communication as means of navigating nutrition and exercise concerns in an online cancer community.* Journal of Medical Internet Research, 2013. **15**(5).
171. Lu, Y., et al., *Health-Related Hot Topic Detection in Online Communities Using Text Clustering.* PLoS ONE, 2013. **8**(2).
172. Lu, Y., et al., *Understanding health care social media use from different stakeholder perspectives: A content analysis of an online health community.* Journal of Medical Internet Research, 2017. **19**(4).
173. Lyles, C.R., et al., *"5 Mins of uncomfyness is better than dealing with cancer 4 a lifetime": An exploratory qualitative analysis of cervical and breast cancer screening dialogue on twitter.* Journal of Cancer Education, 2013. **28**(1): p. 127-133.
174. Mao, J.J., et al., *Online discussion of drug side effects and discontinuation among breast cancer survivors.* Pharmacoepidemiology and Drug Safety, 2013. **22**(3): p. 256-262.
175. Marmor, R.A., et al., *Increase in contralateral prophylactic mastectomy conversation online unrelated to decision-making.* Journal of Surgical Research, 2017. **218**: p. 253-260.
176. Marshall, S.A., et al., *Symptom clusters in women with breast cancer: an analysis of data from social media and a research study.* Quality of Life Research, 2016. **25**(3): p. 547-557.
177. Martinez, B., et al., *Patient Understanding of the Risks and Benefits of Biologic Therapies in Inflammatory Bowel Disease: Insights from a Large-scale Analysis of Social Media Platforms.* Inflammatory Bowel Diseases, 2017. **23**(7): p. 1057-1064.
178. McCosker, A. and R. Darcy, *LIVING WITH CANCER: Affective Labour, Self-expression And The Utility of Blogs.* Information Communication and Society, 2013. **16**(8): p. 1266-1285.
179. Mishra, A., A. Malviya, and S. Aggarwal. *Towards Automatic Pharmacovigilance: Analysing Patient Reviews and Sentiment on Oncological Drugs*. in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. 2015.
180. Mogi, Y., et al., *Social Networking Sites for Peer-Driven Health Communication: Diabetes-Related Communities in Google+.* Diabetology International, 2017. **8**(3): p. 323-327.
181. Myrick, J.G., et al., *#Stupidcancer: Exploring a Typology of Social Support and the Role of Emotional Expression in a Social Media Community.* Health Communication, 2016. **31**(5): p. 596-605.
182. Neal, D.M. and P.J. McKenzie, *Putting the pieces together: Endometriosis blogs, cognitive authority, and collaborative information behavior.* Journal of the Medical Library Association, 2011. **99**(2): p. 127-134.
183. Odh, I., M. Löfving, and K. Klaeson, *Existential challenges in young people living with a cancer diagnosis.* European Journal of Oncology Nursing, 2016. **24**: p. 54-60.
184. Park, H. and M.S. Park, *Cancer information-seeking behaviors and information needs among Korean Americans in the online community.* Journal of Community Health, 2014. **39**(2): p. 213-220.
185. Park, S., et al., *The source and credibility of colorectal cancer information on twitter.* Medicine (United States), 2016. **95**(7): p. e2775.
186. Ploderer, B., et al. *The transition of stroke survivors from hospital to home: Understanding work and design opportunities*. 2017. Association for Computing Machinery.
187. Quinn, E.M., et al., *Who's talking about breast cancer? Analysis of daily breast cancer posts on the internet.* Breast, 2013. **22**(1): p. 24-27.
188. Roccetti, M., et al., *Unleashing the true potential of social networks: confirming infliximab medical trials through Facebook posts.* Network Modeling Analysis in Health Informatics and Bioinformatics, 2016. **5**(1).
189. Rus, H.M. and L.D. Cameron, *Health Communication in Social Media: Message Features Predicting User Engagement on Diabetes-Related Facebook Pages.* Annals of Behavioral Medicine, 2016. **50**(5): p. 678-689.
190. Shah, S.G. and I. Robinson, *Patients' perspectives on self-testing of oral anticoagulation therapy: content analysis of patients' internet blogs.* BMC Health Serv Res, 2011. **11**: p. 25.
191. Sillence, E., *Giving and receiving peer advice in an online breast cancer support group.* Cyberpsychology, Behavior, and Social Networking, 2013. **16**(6): p. 480-485.
192. Sillence, E. and P.K.H. Mo, *Communicating health decisions: An analysis of messages posted to online prostate cancer forums.* Health Expectations, 2014. **17**(2): p. 244-253.

193. Sinnenberg, L., et al., *Twitter as a potential data source for cardiovascular disease research.* JAMA Cardiology, 2016. **1**(9): p. 1032-1036.
194. Stewart Loane, S., C.M. Webster, and S. D'Alessandro, *Identifying Consumer Value Co-created through Social Support within Online Health Communities.* Journal of Macromarketing, 2015. **35**(3): p. 353-367.
195. Strekalova, Y.A. and J.L. Krieger, *Beyond Words: Amplification of Cancer Risk Communication on Social Media.* Journal of Health Communication, 2017. **22**(10): p. 849-857.
196. Strekalova, Y.A. and J.L. Krieger, *A Picture Really is Worth a Thousand Words: Public Engagement with the National Cancer Institute on Social Media.* Journal of Cancer Education, 2017. **32**(1): p. 155-157.
197. Tang, C., et al., *Comment topic evolution on a cancer institution's facebook page.* Applied Clinical Informatics, 2017. **8**(3): p. 854-865.
198. Thackeray, R., et al., *Using Twitter for breast cancer prevention: An analysis of breast cancer awareness month.* BMC Cancer, 2013. **13**.
199. Tsuya, A., et al., *Do cancer patients tweet? Examining the twitter use of cancer patients in Japan.* Journal of Medical Internet Research, 2014. **16**(5).
200. Vaughan Sarrazin, M.S., et al., *Patient perspectives of dabigatran: Analysis of online discussion forums.* Patient, 2014. **7**(1): p. 47-54.
201. Yin, Z., et al., *A scalable framework to detect personal health mentions on Twitter.* Journal of Medical Internet Research, 2015. **17**(6): p. e138.
202. Yin, Z., W. Xie, and B.A. Malin, *Talking About My Care: Detecting Mentions of Hormonal Therapy Adherence Behavior in an Online Breast Cancer Community.* AMIA ... Annual Symposium proceedings. AMIA Symposium, 2017. **2017**: p. 1868-1877.
203. Zhang, J., R. Marmor, and J. Huh, *Towards Supporting Patient Decision-making In Online Diabetes Communities.* AMIA ... Annual Symposium proceedings. AMIA Symposium, 2017. **2017**: p. 1893-1902.
204. Zhang, S., et al., *Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks.* Journal of Biomedical Informatics, 2017. **69**: p. 1-9.
205. *WordItOut - Make a World Cloud*. 2020 [cited 2020; Archived at: http://archive.is/3QAN2]. Available from: https://worditout.com/word-cloud/create.
206. Govind. *THE INTERNET IS THE NEW GEOGRAPHY: NATION-STATES BEWARE, NETIZENS ARE THE CITIZENS OF THE FUTURE*. 2016 [cited 2020; Archived at: http://archive.is/giZ0a]. Available from: https://www.firstpost.com/tech/news-analysis/the-internet-is-the-new-geography-nation-states-beware-netizens-are-the-citizens-of-the-future-3694889.html.
207. Gillum, L.A., et al., *NIH Disease Funding Levels and Burden of Disease.(Research Article).* PLoS ONE, 2011. **6**(2): p. e16837.
208. *What is cancer?* [cited 2020; Archived at: http://archive.is/xmZ4N]. Available from: https://www.cancerresearchuk.org/about-cancer/what-is-cancer.
209. WHO. *Cancer*. 2018 [cited 2020; Archived at: http://archive.is/CGej5]. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.
210. Gruzd, A. *Current State of Social Media Research: From Practice to Theory #SMSociety15*. 2015; Archived at: http://www.webcitation.org/6q86pewm0]. Available from: http://socialmedialab.ca/2015/current-state-of-social-media-research-from-practice-to-theory-part-1/.
211. Knapp, B., et al., *Ten simple rules for a successful cross-disciplinary collaboration.* PLoS computational biology, 2015. **11**(4): p. e1004214-e1004214.
212. Willyard, C. *The Agony and Ecstasy of Cross-disciplinary Collaboration*. 2013 [cited 2020; Archived at: http://archive.is/sy8qd]. Available from: https://www.sciencemag.org/careers/2013/08/agony-and-ecstasy-cross-disciplinary-collaboration.
213. Urquhart, R., et al., *Cross-disciplinary research in cancer: an opportunity to narrow the knowledge-practice gap.* Current oncology (Toronto, Ont.), 2013. **20**(6): p. e512.
214. Koch, S., *Healthy ageing supported by technology – a cross-disciplinary research challenge.* Informatics for Health and Social Care, 2010. **35**(3-4): p. 81-91.
215. Xu, R. and N. Ghaffarzadegan, *Neuroscience bridging scientific disciplines in health: Who builds the bridge, who pays for it?* Scientometrics, 2018. **117**(2): p. 1183-1204.

216. Börner, K., et al., *A multi-level systems perspective for the science of team science.* Science Translational Medicine, 2010. **2**(49).
217. Benner, M. and U. Sandström, *Institutionalizing the triple helix: research funding and norms in the academic system.* Research Policy, 2000. **29**(2): p. 291-301.
218. Stirling, A. *Disciplinary dilemma: working across research silos is harder than it looks*. 2014 [cited 2020; Archived at: http://archive.is/pgT7p]. Available from: https://www.theguardian.com/science/political-science/2014/jun/11/science-policy-research-silos-interdisciplinarity.
219. Aagaard-Hansen, J., *The Challenges of Cross-disciplinary Research.* Social Epistemology, 2007. **21**(4): p. 425-438.
220. JMIR. *Editorial Policies*. [cited 2020 25 April]; Archived at: http://archive.is/UyKsd]. Available from: https://www.jmir.org/about/editorialPolicies - custom10.
221. IEEE. *About IEEE*. Archived at: http://archive.is/lar3o]. Available from: https://www.ieee.org/about/index.html.
222. AoIR. *Internet Research: Ethical Guidelines 3.0* 2019 [cited 2020; Archived at: http://archive.is/tBwOo]. Available from: https://aoir.org/reports/ethics3.pdf.
223. IEEE. *IEEE Authors Ethics Guidelines*. [cited 2020; Archived at: http://archive.is/pY4AE/image]. Available from: https://journals.ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Author-Ethics-Guidelines.pdf.
224. Confessore, N. *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*. 2018; Archived at: http://archive.is/RZ8wX]. Available from: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.
225. *Social Media Fact Sheet*. 2019; Archived at: http://archive.is/otreS]. Available from: https://www.pewresearch.org/internet/fact-sheet/social-media/ - which-social-media-platforms-are-most-popular.
226. *Twitter data for academic research*. [cited 2020; Archived at: http://archive.is/gzl1j]. Available from: https://developer.twitter.com/en/use-cases/academic-researchers.
227. *Why the Healthcare Hashtag Project?* ; Archived at: http://archive.is/hGwPQ]. Available from: https://www.symplur.com/healthcare-hashtags/.
228. *Research Areas*. Archived at: http://archive.is/wip/ROrlV]. Available from: https://research.fb.com/research-areas/.
229. *Facebook for developers*. Archived at: http://archive.is/pBbSS/image]. Available from: https://developers.facebook.com.
230. *About WebMD*. [cited 2020; Archived at: http://archive.is/4icHw]. Available from: https://www.webmd.com/about-webmd-policies/default.htm.
231. *About Breastcancer.org*. [cited 2020; Archived at: http://archive.is/K3j9X]. Available from: https://www.breastcancer.org/about_us.
232. *About Diabetes Daily*. [cited 2020; Archived at: http://archive.is/10RWj]. Available from: https://www.diabetesdaily.com/about/.
233. Pushpam, C.A. and J.G. Jayanthi, *Over view on Data Mining in Social Media.* International Journal of Computer Sciences and Engineering, 2017. **5**(11): p. 147-157.
234. Cortizo, J.C., F.M. Carrero, and J.M. Gómez, *Introduction to the Special Issue: Mining Social Media.* International Journal of Electronic Commerce, 2011. **15**(3): p. 5-8.
235. Zafarani, R., M.A. Abbasi, and H. Liu, *Social media mining: An introduction*. Vol. 9781107018853. 2014. 1-320.
236. Paul, M.J., et al., *SOCIAL MEDIA MINING FOR PUBLIC HEALTH MONITORING AND SURVEILLANCE*, in *Biocomputing 2016*. 2015, WORLD SCIENTIFIC. p. 468-479.
237. Duarte, N., E. Llanso, and A. Loup, *Mixed Messages? The Limits of Automated Social Media Content Analysis*, in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, A.F. Sorelle and W. Christo, Editors. 2018, PMLR: Proceedings of Machine Learning Research. p. 106--106.
238. *Scimago Journal and Country Rank*. [cited 2020; Archived at: http://archive.is/wip/RtBgC]. Available from: https://www.scimagojr.com/countryrank.php.
239. McCarthy, N. *These countries spend the most on university research*. 2019 [cited 2020; Archived at: http://archive.is/mLrBq]. Available from: https://www.weforum.org/agenda/2019/11/us-government-funded-university-research-slips-rankings/.
240. Foote, R.D.A.a.C. *U.S. Funding for University Research Continues to Slide*. 2019 [cited Archived at: http://archive.is/hnThe 2020]; Available from: https://itif.org/publications/2019/10/21/us-funding-university-research-continues-slide.

241. *Countries arranged by Number of Universities in Top Ranks*. 2020; Archived at: http://archive.is/hwm4R]. Available from: https://www.webometrics.info/en/node/54.
242. *QS World University Rankings by Subject*. 2020; Archived at: http://archive.is/qBA6K]. Available from: https://www.topuniversities.com/subject-rankings/2020.
243. Linda, B. and M. Ian, *Evaluating University Research Performance Using Metrics.* European Political Science, 2011. **10**(1): p. 44.
244. Pinfield, S., *Paying for open access? Institutional funding streams and OA publication charges.* Learned Publishing, 2010. **23**(1): p. 39-52.
245. WHO. *Noncommunicable diseases country profiles 2018*. 2018 [cited 2020; Archived at: http://archive.is/SLspt]. Available from: https://www.who.int/nmh/countries/en/.
246. Clement, J. *Number of global social network users 2010-2023*. 2020 [cited 2020; Archived at: http://archive.is/z710C]. Available from: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.
247. *Active social network penetration in selected countries as of January 2020*. 2020 [cited 2020; Archived at: http://archive.is/FMxV4]. Available from: https://www.statista.com/statistics/282846/regular-social-networking-usage-penetration-worldwide-by-country/.
248. Taylor, H. *If social networks were countries, which would they be?* 2016 [cited 2020; Archived at: http://archive.is/2I6bk]. Available from: https://www.weforum.org/agenda/2016/04/facebook-is-bigger-than-the-worlds-largest-country/.
249. Clement, J. *Social network users in selected countries in 2018 and 2023*. 2019 [cited 2020; Archived at: http://archive.is/cQZdU]. Available from: https://www.statista.com/statistics/278341/number-of-social-network-users-in-selected-countries/.
250. van Dijk, J.A.G.M., *Digital divide research, achievements and shortcomings.* Poetics, 2006. **34**(4-5): p. 221-235.
251. *Digital inclusion of all*. 2019; Archived at: http://archive.is/vkal2]. Available from: https://www.itu.int/en/mediacentre/backgrounders/Pages/digital-inclusion-of-all.aspx.
252. Blank, G. and C. Lutz, *Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram.* American Behavioral Scientist, 2017. **61**(7): p. 741-756.
253. Kennedy, H. and G. Moss, *Known or knowing publics? Social media data mining and the question of public agency.* Big Data & Society, 2015. **2**(2): p. 2053951715611145.
254. Blank, G., *The Digital Divide Among Twitter Users and Its Implications for Social Research.* Social Science Computer Review, 2016. **35**(6): p. 679-697.
255. Taylor, J. and C. Pagliari, *Comprehensive scoping review of health research using social media data.* BMJ Open, 2018. **8**(12): p. e022931.
256. WHO. *World Health Statistics 2016 data visualizations dashboard: Monitoring health for SDGs*. 2016; Archived at: http://www.webcitation.org/6q886MFDK]. Available from: http://apps.who.int/gho/data/node.sdg.tp-1?lang=en.
257. *Addressing NCD Co-Morbidities:Shared Opportunities for Action*. [cited 2020; Archived at: http://archive.is/YIxJ3]. Available from: https://ncdalliance.org/sites/default/files/resource_files/Brochure_Osteoporosis%26NCDs_WEB.pdf.
258. O'Connor, A.M., et al., *Still moving toward automation of the systematic review process: a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR).* Systematic Reviews, 2019. **8**(1): p. 57.
259. Tsafnat, G., et al., *The automation of systematic reviews.* BMJ : British Medical Journal, 2013. **346**: p. f139.
260. Tsafnat, G., et al., *Systematic review automation technologies.* Systematic Reviews, 2014. **3**(1): p. 74.
261. BPS. *Conducting Research on the Internet - Guidelines for ethical practice in psychological research online*. 2012; Archived at: http://www.webcitation.org/6NiWqxCb6]. Available from: http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf.
262. Markham, A. and E. Buchanan. *Ethical Decision-Making and Internet Research*. 2012; Archived at: http://www.webcitation.org/6NiWZFLIg]. Available from: http://www.aoir.org/reports/ethics2.pdf.

263. *Longitudinal vs cross-sectional studies*. [cited 2020; Archived at: http://archive.vn/LWxWV]. Available from: https://learning.closer.ac.uk/introduction/types-of-longitudinal-research/longitudinal-versus-cross-sectional-studies/.
264. *Cross-Sectional Study: Definition, Advantages, Disadvantages & Example*. [cited 2020; Archived at: http://archive.vn/LUMtP]. Available from: https://study.com/academy/lesson/cross-sectional-study-definition-advantages-disadvantages-example.html.
265. *History and Development of PRISMA*. 2015 [cited 2020; Archived at: http://archive.vn/QrkZi]. Available from: http://prisma-statement.org/PRISMAStatement/HistoryAndDevelopment.
266. Page, M.J., et al., *Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study.* PLOS Medicine, 2016. **13**(5): p. e1002028.
267. Collaboration, C. *Glossary of Terms in The Cochrane Collaboration Version 4.2.5 Updated May 2005*. Archived at: http://www.webcitation.org/6NfTcvMR4]. Available from: http://www.cochrane.org/sites/default/files/uploads/glossary.pdf.
268. Grant, M.J. and A. Booth, *A typology of reviews: an analysis of 14 review types and associated methodologies.* Health Information & Libraries Journal, 2009. **26**(2): p. 91-108.
269. Pickering, C. and J. Byrne, *The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers.* Higher Education Research & Development, 2014. **33**(3): p. 534-548.
270. *Better evidence for a better world*. Archived at: https://archive.is/Dd36K]. Available from: https://campbellcollaboration.org.
271. Resnik, D.B. *What Is Ethics in Research & Why Is It Important?* 2015 [cited 2020; Archived at: http://archive.is/zKpgl]. Available from: https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm.
272. Munson, S.A., et al., *Sociotechnical challenges and progress in using social media for health.* Journal Of Medical Internet Research, 2013. **15**(10): p. e226-e226.
273. David, M.B., *Internet research: privacy, ethics and alienation: an open source approach.* Internet Research, 2004. **14**(4): p. 323-332.
274. Eysenbach, G. and J.E. Till, *Ethical issues in qualitative research on internet communities.* BMJ, 2001. **323**(7321): p. 1103.
275. Koene, A. and S. Adolphs. *Ethics considerations for Corpus Linguistic studies using internet resources*. 2015; Archived at: http://www.webcitation.org/6qcb23iJo]. Available from: http://casma.wp.horizon.ac.uk/wp-content/uploads/2015/04/CL2015-CorpusLinguisticsEthics_KoeneAdolphs.pdf
276. McKee, R., *Ethical issues in using social media for health and health care research.* Health Policy, 2013. **110**(2-3): p. 298-301.
277. Nunan, D. and B. Yenicioglu, *Informed, uninformed and participative consent in social media research.(Report).* International Journal of Market Research, 2013. **55**(6): p. 791.
278. Orton-Johnson, K., *Ethics in Online Research; Evaluating the ESRC Framework for Research Ethics Categorisation of Risk.* Sociological Research Online, 2010. **15**(4): p. 13.
279. Vayena, E., A. Mastroianni, and J. Kahn, *Ethical Issues in Health Research With Novel Online Sources.* American Journal of Public Health, 2012. **102**(12): p. 2225-2230.
280. Jouhki, J., et al., *Facebook's Emotional Contagion Experiment as a Challenge to Research Ethics.* Media and Communication, 2016. **4**(4).
281. Pells, R. *Facebook research targeted insecure youth, leaked documents show*. 2017; Archived at: http://www.webcitation.org/6qcbOx82V]. Available from: http://www.independent.co.uk/news/media/facebook-leaked-documents-research-targeted-insecure-youth-teenagers-vulnerable-moods-advertising-a7711551.html
282. Fromm, J. *New Study Finds Social Media Shapes Millennial Political Involvement and Engagement*. Forbes 2016; Archived at: http://www.webcitation.org/6qcYUC8jd]. Available from: https://www.forbes.com/sites/jefffromm/2016/06/22/new-study-finds-social-media-shapes-millennial-political-involvement-and-engagement/ - 5e93a7302618.
283. Arthur, C. *2010: The first social media election*. 2010; Archived at: http://www.webcitation.org/6qcYqDGMr]. Available from: https://www.theguardian.com/media/2010/apr/30/social-media-election-2010.
284. ESRC. *ESRC Framework for Research Ethics*. 2012; Archived at: http://www.webcitation.org/6NiWjrith]. Available from: http://www.esrc.ac.uk/_images/Framework-for-Research-Ethics_tcm8-4586.pdf.

285. Whiteman, N., *Undoing ethics [electronic resource] : rethinking practice in online research / Natasha Whiteman*. 2012: New York : Springer, [2012], ©2012.

286. Gov.UK. *Data protection*. Archived at: http://www.webcitation.org/6sZLF6OQ9]. Available from: https://www.gov.uk/data-protection/the-data-protection-act.

287. homepage, R. *Homepage*. Archived at: http://www.webcitation.org/6cMlqKpHU]. Available from: http://www.rcuk.ac.uk.

288. RCUK. [cited 2015; Archived at: http://www.webcitation.org/6WhQyTyMS]. Available from: http://www.rcuk.ac.uk.

289. AHRC. *Home page*. [cited 2015; Archived at: http://www.webcitation.org/6WhTE10dY]. Available from: http://www.ahrc.ac.uk/Pages/Home.aspx.

290. BBSRC. *Home page*. [cited 2015; Archived at: http://www.webcitation.org/6WhUhYwCe]. Available from: http://www.bbsrc.ac.uk/home/home.aspx.

291. EPSRC. *Home page*. [cited 2015; Archived at: http://www.webcitation.org/6WhUvsgXo]. Available from: http://www.epsrc.ac.uk.

292. ESRC. *Home page*. [cited 2015; Archived at: http://www.webcitation.org/6WhVuaPvV]. Available from: http://www.esrc.ac.uk.

293. MRC. *Home page*. [cited 2015; Archived at: http://www.webcitation.org/6WhWyd42J]. Available from: http://www.mrc.ac.uk.

294. NERC. *Home page*. [cited 2015; Archived at: http://www.webcitation.org/6WhXaCWcL]. Available from: http://www.nerc.ac.uk.

295. STFC. *Home page*. Archived at: http://www.webcitation.org/6WhYNsN9I]. Available from: https://www.stfc.ac.uk/home.aspx.

296. RCUK. *RCUK Policy and Guidelines on Governance of Good Research Conduct*. 2013 [cited 2015; Archived at: http://www.webcitation.org/6WhbA91Zi]. Available from: http://www.rcuk.ac.uk/RCUK-prod/assets/documents/reviews/grc/RCUKPolicyandGuidelinesonGovernanceofGoodResearchPracticeFebruary2013.pdf.

297. AHRC. *Research Funding Guide*. 2016; Archived at: http://www.webcitation.org/6qQMS16WS]. Available from: http://www.ahrc.ac.uk/documents/guides/research-funding-guide/.

298. BBSRC. *BBSRC Research Grants The Guide*. 2017; Archived at: http://www.webcitation.org/6qQMmj0ov]. Available from: http://www.bbsrc.ac.uk/documents/grants-guide/.

299. EPSRC. *Framework for Responsible Innovation*. 2013; Archived at: http://www.webcitation.org/6WhVayhjf]. Available from: http://www.epsrc.ac.uk/research/framework/.

300. MRC. *MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies*. 2012; Archived at: http://www.webcitation.org/6XJbxQoUX]. Available from: http://www.mrc.ac.uk/news-events/publications/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/.

301. MRC. *Personal Information in Medical Research*. 2000; Archived at: http://www.webcitation.org/6XJcNsQx9]. Available from: http://www.mrc.ac.uk/documents/pdf/personal-information-in-medical-research/.

302. MRC. *Good research practice*. 2012; Archived at: http://www.webcitation.org/6XJcio3CY]. Available from: http://www.mrc.ac.uk/news-events/publications/good-research-practice-principles-and-guidelines/.

303. NERC. *Ethics Policy*. 2015 [cited 2015; Archived at: http://www.webcitation.org/6WhY6SCE7]. Available from: http://www.nerc.ac.uk/about/policy/policies/nerc-ethics-policy.pdf.

304. STFC. *Public Engagement with Science and Technology Strategic Plan 2013-2016*. 2013 [cited 2015; Archived at: http://www.webcitation.org/6WhYtNpfM]. Available from: http://www.stfc.ac.uk/files/1328/1328_res_2.pdf.

305. ESRC. *Framework for research ethics*. 2015 [cited 2015; Archived at: http://www.webcitation.org/6WhWkKPQN]. Available from: http://www.esrc.ac.uk/_images/framework-for-research-ethics_tcm8-33470.pdf.

306. BPS. *Ethics Guidelines for Internet-mediated Research*. 2017 [cited 2020; Archived at: http://archive.is/Ouwvm]. Available from: https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy - Files/Ethics Guidelines for Internet-mediated Research %282017%29.pdf.

307. SOCIAM. *SOCIAM: The Theory and Practice of Social Machines*. Archived at: http://www.webcitation.org/6cXbwBjyN]. Available from: http://sociam.org.
308. McKee, H.A., *Policy Matters Now and in the Future: Net Neutrality, Corporate Data Mining, and Government Surveillance.* Computers and Composition, 2011. **28**(4): p. 276-291.
309. Mackey, T.K., et al., *A call for a moratorium on the .health generic top-level domain: preventing the commercialization and exclusive control of online health information.* Globalization and Health, 2014. **10**(1): p. 62.
310. Kiss, J. *An online Magna Carta: Berners-Lee calls for bill of rights for web*. 2014; Archived at: http://www.webcitation.org/6cXbhHWo1]. Available from: http://www.theguardian.com/technology/2014/mar/12/online-magna-carta-berners-lee-web.
311. O'Sullivan, B., *Towards a Magna Carta for Data*. 2017, Royal Irish Academy.
312. Knapton, S. *Health records of every NHS patient to be shared in vast database*. 2014; Archived at: http://www.webcitation.org/6qcbv3pMG]. Available from: http://www.telegraph.co.uk/news/10565160/Health-records-of-every-NHS-patient-to-be-shared-in-vast-database.html.
313. Boseley, S. *NHS to scrap single database of patients' medical details*. 2016; Archived at: http://www.webcitation.org/6qccG78Zw]. Available from: https://www.theguardian.com/technology/2016/jul/06/nhs-to-scrap-single-database-of-patients-medical-details.
314. Wakefield, J. *Google DeepMind's NHS deal under scrutiny*. 2017; Archived at: http://www.webcitation.org/6qccqJ1jz]. Available from: http://www.bbc.com/news/technology-39301901.
315. Guardian, N.D. *National Data Guardian (NDG) statement on government response to the NDG Review*. 2017; Archived at: http://www.webcitation.org/6t9wBQUgI]. Available from: https://www.gov.uk/government/news/national-data-guardian-ndg-statement-on-government-response-to-the-ndg-review.
316. Convery, I. and D. Cox, *A review of research ethics in internet-based research.* Practitioner Research in Higher Education, 2012. **6**(1).
317. NCCPE. *Ethical guidelines web resources*. Archived at: http://www.webcitation.org/6clBIHhfW]. Available from: http://www.publicengagement.ac.uk/work-with-us/completed-projects/ethics-cbpr/resources/ethical-guidelines-web-resources.
318. Bauchner, H. and R. Vinci, *What have we learnt from the Alder Hey affair? That monitoring physicians' performance is necessary to ensure good practice.* BMJ (Clinical research ed.), 2001. **322**(7282): p. 309-310.
319. Pasquale, F., *The black box society : the secret algorithms that control money and information*. 2015: Cambridge : Harvard University Press.
320. Metcalf, J., E.F. Keller, and D. Boyd. *Perspectives on Big Data, Ethics, and Society*. 2017; Archived at: ttp://http://www.webcitation.org/6t9ySFWQF]. Available from: http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/.
321. Evans, H., S. Ginnis, and J. Bartlett. *#SocialEthics a guide to embedding ethics in social media research*. 2015; Archived at: http://www.webcitation.org/6t9xdn1QT]. Available from: https://www.ipsos.com/sites/default/files/migrations/en-uk/files/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf.
322. OECD, *Research Ethics and New Forms of Data for Social and Economic Research*. 2016: OECD Publishing.
323. Bishop, L. *Big data and data sharing: Ethical issues*. 2017; Archived at: http://www.webcitation.org/6t9ynVdaG]. Available from: https://www.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf.
324. SocDM. *Society of Data Miners: Towards a Professional Code of Ethics for Data Mining*. 2017; Archived at:]. Available from: https://www.turing.ac.uk/events/socdm-towards-professional-code-ethics-data-mining-part-2/.
325. Aitken, M., et al., *Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies.* BMC Med Ethics, 2016. **17**(1): p. 73.
326. Williams, M.L., P. Burnap, and L. Sloan, *Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation.* Sociology, 2017: p. 0038038517708140.
327. Commission, E. *Options for Strengthening Responsible Research and Innovation - Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation*.

2013; Archived at: http://www.webcitation.org/6tA1IklBr]. Available from: http://ec.europa.eu/research/science-society/document_library/pdf_06/options-for-strengthening_en.pdf
328. Parliament, E., *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. 2016.
329. Hookway, N., *'Entering the blogosphere': some strategies for using blogs in social research.* Qualitative Research, 2008. **8**(1): p. 91-113.
330. R Lyles, C., et al., *"5 Mins of uncomfyness is better than dealing with cancer 4 a lifetime": an exploratory qualitative analysis of cervical and breast cancer screening dialogue on twitter.* Journal of Cancer Education, 2013. **28**(1): p. 127-133.
331. Himelboim, I. and J.Y. Han, *Cancer Talk on Twitter: Community Structure and Information Sources in Breast and Prostate Cancer Social Networks.* Journal Of Health Communication, 2013.
332. Marton, C., *Consumer health 2.0 in Canada: tweeting about cancer.* Journal of the Canadian Health Libraries Association (JCHLA), 2012. **33**(2): p. 98-102.
333. Robillard, J.M., et al., *Aging 2.0: health information about dementia on Twitter.* Plos One, 2013. **8**(7): p. e69861-e69861.
334. McNeil, K., P.M. Brna, and K.E. Gordon, *Epilepsy in the Twitter era: a need to re-tweet the way we think about seizures.* Epilepsy & Behavior: E&B, 2012. **23**(2): p. 127-130.
335. Sirven, J.I., *It's time for an epilepsy Twitter revolution.* Epilepsy & Behavior: E&B, 2012. **23**(2): p. 97-97.
336. Love, B., et al., *Twitter as a source of vaccination information: Content drivers and what they are saying.* American Journal of Infection Control, 2013. **41**(6): p. 568-570.
337. Gold, J., et al., *A systematic examination of the use of online social networking sites for sexual health promotion.* BMC Public Health, 2011. **11**: p. 583-583.
338. Burton, S., A. Dadich, and A. Soboleva, *Competing Voices: Marketing and Counter-Marketing Alcohol on Twitter.* Journal of Nonprofit & Public Sector Marketing, 2013. **25**(2): p. 186-209.
339. Chan, K., S. Ho, and T. Lam, *Infodemiology of alcohol use in Hong Kong mentioned on blogs: infoveillance study.* Journal Of Medical Internet Research, 2013. **15**(9): p. e192-e192.
340. Prochaska, J.J., et al., *Twitter=quitter? An analysis of Twitter quit smoking social networks.* Tobacco Control, 2012. **21**(4): p. 447-449.
341. Myslín, M., et al., *Using twitter to examine smoking behavior and perceptions of emerging tobacco products.* Journal Of Medical Internet Research, 2013. **15**(8): p. e174-e174.
342. St Louis, C. and G. Zorlu, *Can Twitter predict disease outbreaks?* British Medical Journal, 2012. **344**(7861): p. 24.
343. Denecke, K., et al., *How to exploit twitter for public health monitoring?* Methods of Information in Medicine, 2013. **52**(4): p. 326-339.
344. Collier, N., N.T. Son, and N.M. Nguyen, *OMG U got flu? Analysis of shared health messages for bio-surveillance.* 2011.
345. Chew, C. and G. Eysenbach, *Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak.* Plos One, 2010. **5**(11): p. e14118-e14118.
346. *Tracking public health via Twitter.(DISEASE WATCH: IN THE NEWS)(Brief article).* Nature Reviews Microbiology, 2011(9): p. 634.
347. Dwoskin, E. *What Twitter Can Tell Us About Ebola*. 2014; Archived at: http://www.webcitation.org/6V6atSMUz]. Available from: http://blogs.wsj.com/digits/2014/11/25/what-twitter-can-tell-us-about-ebola/.
348. Gui, X., et al., *Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak.* AMIA ... Annual Symposium proceedings. AMIA Symposium, 2017. **2017**: p. 820-829.
349. WHO. Archived at: http://www.webcitation.org/6Ohk9aWEI]. Available from: http://www.who.int/mediacentre/factsheets/fs312/en/.
350. IDF. *Diabetes Atlas 2010*. 2000; Archived at: http://www.webcitation.org/6RtIqPenT]. Available from: http://www.idf.org/sites/default/files/IDF Diabetes Atlas-2000 %281st edition%29_0.pdf.
351. IDF. *Diabetes Atlas Second Edition*. 2003; Archived at: http://www.webcitation.org/6RtJHtS1Z]. Available from: http://www.idf.org/sites/default/files/IDF_Diabetes_Atlas_2ndEd.pdf.

352. IDF. *Diabetes Atlas Third Edition*. 2006; Archived at: http://www.webcitation.org/6RtJPKrYb]. Available from: http://www.idf.org/sites/default/files/da5/Diabetes Atlas 3rd edition.pdf.
353. IDF. *Diabetes Atlas Fifth Edition*. 2012; Archived at: http://www.webcitation.org/6RtJeEwWZ]. Available from: http://www.idf.org/sites/default/files/5E_IDFAtlasPoster_2012_EN.pdf.
354. IDF. *Diabetes Atlas Sixth Edition*. 2013; Archived at: http://www.webcitation.org/6RtJqC9tf]. Available from: http://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf.
355. Duggan, M. and A. Smith. *Social Media Update 2013*. 2013; Archived at: http://www.webcitation.org/6NiXdDQl0]. Available from: http://pewinternet.org/~/media//Files/Reports/2013/Social Networking 2013_PDF.pdf.
356. Fox, S. *Health Topics*. 2011; Archived at: http://www.webcitation.org/6NiXCoo2q]. Available from: http://www.pewinternet.org/~/media//Files/Reports/2011/PIP_Health_Topics.pdf.
357. *Twitter Usage Statistics*. [cited 2020; Archived at: https://archive.is/cYeU7]. Available from: https://www.internetlivestats.com/twitter-statistics/.
358. Protalinski, E. *Twitter sees 218m monthly active users, 163.5m monthly mobile users, 100m daily users, and 500m tweets per day*. 2013; Archived at: https://archive.is/KhSk7]. Available from: https://thenextweb.com/twitter/2013/10/03/twitter-says-it-sees-215-million-monthly-active-users-100-million-daily-users-and-500-million-tweets-per-day/.
359. Twitter. *GET statuses / sample*. Archived at: http://www.webcitation.org/6W0vz45mF]. Available from: https://dev.twitter.com/streaming/reference/get/statuses/sample.
360. Harden, A. *Mixed-Methods Systematic Reviews: Integrating Quantitative and Qualitative Findings*. Focus: Technical Brief No. 25 2010; Archived at: http://www.webcitation.org/6NiWR4bUF]. Available from: http://www.ktdrr.org/ktlibrary/articles_pubs/ncddrwork/focus/focus25/Focus25.pdf.
361. Bales, R.F., *Interaction process analysis : a method for the study of small groups / by Robert F. Bales*. 1951: Cambridge, Mass. : Addison-Wesley Press.
362. Roter, D. and S. Larson, *The Roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions.* PATIENT EDUCATION AND COUNSELING, 2002. **46**(4): p. 243-251.
363. Benne, K.D. and P. Sheats, *Functional Roles of Group Members.* Journal of Social Issues, 1948. **4**(2): p. 41-49.
364. Bartholomew, D.J., *The analysis and interpretation of multivariate data for social scientists / David J. Bartholomew [and others]*. Chapman & Hall/CRC texts in statistical science series. 2002: Boca Raton, Fla. ; London : Chapman & Hall/CRC, [2002], ©2002.
365. Matsumoto, K., et al., *Slang feature extraction by analysing topic change on social media.* CAAI Transactions on Intelligence Technology, 2019. **4**(1): p. 64-71.
366. Donelle, L. and R.G. Booth, *Health tweets: an exploration of health promotion on twitter.* Online Journal Of Issues In Nursing, 2012. **17**(3): p. 4-4.
367. Harris, J.K., et al., *Communication About Childhood Obesity on Twitter.* American Journal of Public Health, 2014. **104**(7): p. e62-e69.
368. Mackert, M., et al., *Using twitter for prenatal health promotion: encouraging a multivitamin habit among college-aged females.* Studies in Health Technology & Informatics, 2012. **182**: p. 93-103.
369. Ramanadhan, S., et al., *Social media use by community-based organizations conducting health promotion: a content analysis.* BMC PUBLIC HEALTH, 2013. **13**.
370. Rui, J.R., C. Yixin, and A. Damiano, *Health Organizations Providing and Seeking Social Support: A Twitter-Based Content Analysis (English).* Cyberpsychology, behavior and social networking (Print), 2013. **16**(9): p. 669-673.
371. Thackeray, R., et al., *Adoption and use of social media among public health departments.* BMC Public Health, 2012. **12**: p. 242-242.
372. Turner-McGrievy, G.M. and D.F. Tate, *Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention.* Translational Behavioral Medicine, 2013. **3**(3): p. 287-294.
373. Rui, H. and A. Whinston, *Information or attention? An empirical study of user contribution on Twitter.* Information Systems and e-Business Management, 2012. **10**(3): p. 309-324.
374. Zhang, Y., et al., *Facebook as a platform for health information and communication: a case study of a diabetes group.* J Med Syst, 2013. **37**(3): p. 9942.

375. CDC. *Successes and Opportunities for Population-Based Prevention and Control At A Glance 2011*. 2011; Archived at: http://www.webcitation.org/6T1PZfpp0]. Available from: http://www.cdc.gov/chronicdisease/resources/publications/aag/ddt.htm.

376. Diabetes.org.uk. *Diabetes in the UK 2010: Key statistics on diabetes*. 2010; Archived at: http://www.webcitation.org/6T1PorClN]. Available from: http://www.diabetes.org.uk/documents/reports/diabetes_in_the_uk_2010.pdf.

377. Clement, J. *Twitter: distribution of global audiences 2020, by age group*. 2020 [cited 2020; Archived at: https://archive.is/3DOK4]. Available from: https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/.

378. Akoumianakis, D., et al. *Networking Tactics for Online eHealth Campaigns and Intervention Studies: Preliminary Results from a Case Study on Nutritional Support for Cancer Patients* . in *Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on*. 2012

379. Tsuya, A., et al., *Do Cancer Patients Tweet? Examining the Twitter Use of Cancer Patients in Japan.* J Med Internet Res, 2014. **16**(5): p. e137.

380. Mano, R.S., *Social media, social causes, giving behavior and money contributions.* Computers in Human Behavior, 2014. **31**: p. 287-293.

381. Moore, C., *Choose the Social-Media Networks That Fit Your Charity's Goals.* Chronicle of Philanthropy, 2014(06).

382. Paulin, M., et al., *Motivating millennials to engage in charitable causes through social media.* Journal of Service Management, 2014: p. 334-348.

383. Johnson, S.A., *"Maternal Devices", Social Media and the Self-Management of Pregnancy, Mothering and Child Health.* Societies (2075-4698), 2014. **4**(2): p. 330.

384. McDaniel, B., S. Coyne, and E. Holmes, *New Mothers and Media Use: Associations Between Blogging, Social Networking, and Maternal Well-Being.* Maternal & Child Health Journal, 2012. **16**(7): p. 1509-1517.

385. Marsden, D., et al., *The role of technology in learning disability nursing. (Cover story).* Learning Disability Practice, 2012. **15**(3): p. 28-30.

386. Nikolaou, I., *Social Networking Web Sites in Job Search and Employee Recruitment.* INTERNATIONAL JOURNAL OF SELECTION AND ASSESSMENT, 2014. **22**(2): p. 179-189.

387. Van de Belt, T.H., et al., *Use of social media by Western European hospitals: longitudinal study.* Journal Of Medical Internet Research, 2012. **14**(3): p. e61-e61.

388. Allen, H.G., et al., *Social Media Release Increases Dissemination of Original Articles in the Clinical Pain Sciences.* PLoS ONE, 2013. **8**(7): p. 1.

389. Gardois, P., et al., *Implementation of Web 2.0 services in academic, medical and research libraries: a scoping review.* Health Information And Libraries Journal, 2012. **29**(2): p. 90-109.

390. Verhagen, E., C. Bower, and K.M. Khan, *How BJSM embraces the power of social media to disseminate research*, in *British Journal of Sports Medicine*. 2014. p. 680.

391. Bornmann, L., *Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics.* Journal of Informetrics, 2014. **8**(4): p. 895-903.

392. NHS. *Type 2 diabetes*. Archived at: http://www.webcitation.org/6T1YVWtUn]. Available from: http://www.nhs.uk/conditions/Diabetes-type2/Pages/Introduction.aspx.

393. Morstatter, F., et al., *Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose.* 2013.

394. Hall, A.K., et al., *The Digital Health Divide: Evaluating Online Health Information Access and Use Among Older Adults.* Health Education & Behavior: The Official Publication Of The Society For Public Health Education, 2014.

395. Heaivilin, N., et al., *Public Health Surveillance of Dental Pain via Twitter.* JOURNAL OF DENTAL RESEARCH, 2011. **90**(9): p. 1047-1051.

396. Frost, J. and M. Massagli, *PatientsLikeMe the case for a data-centered patient community and how ALS patients use the community to inform treatment decisions and manage pulmonary health.* Chronic Respiratory Disease, 2009. **6**(4): p. 225-229.

397. Abdulla, S. *What percentage of Twitter users are fake?* 2017; Archived at: http://www.webcitation.org/753giVZOP]. Available from: https://www.quora.com/What-percentage-of-Twitter-users-are-fake.

398. Symplur. *Disease Hashtags*. Archived at: http://www.webcitation.org/6UTOvbi1D]. Available from: http://www.symplur.com/healthcare-hashtags/diseases/.

399. *Type 1 Diabetes Mellitus*. 2018 [cited 2020; Archived at: http://archive.is/2qdrK]. Available from: https://www.health.harvard.edu/a_to_z/type-1-diabetes-mellitus-a-to-z.

400. *Type 1 Diabetes*. Archived at: http://archive.is/VZZ6R]. Available from: https://www.facebook.com/T1Diabetes.
401. *Type 1 Diabetes*. [cited 2015; Archived at: http://archive.is/32jfc]. Available from: http://www.diabetes.co.uk/forum/category/type-1-diabetes.19/.
402. Kamps, H.J. *Who are Twitter's verified users?* 2015; Archived at: http://www.webcitation.org/753fjjQCa]. Available from: https://medium.com/@Haje/who-are-twitter-s-verified-users-af976fc1b032.
403. WHO. *Cancer Fact Sheet*. 2018; Archived at: http://www.webcitation.org/6yLxyCMD1]. Available from: http://www.who.int/mediacentre/factsheets/fs297/en/.
404. lungcancer.org. *Types and Staging of Lung Cancer*. 2018; Archived at: http://www.webcitation.org/753LJ6cNQ]. Available from: https://www.lungcancer.org/find_information/publications/163-lung_cancer_101/268-types_and_staging.
405. Macmillan.org.uk. *A practical guide to understanding cancer - Lung Cancer*. 2017; Archived at: http://www.webcitation.org/753LZERsq]. Available from: http://be.macmillan.org.uk/Downloads/CancerInformation/CancerTypes/Mac11632LungE16Low-res-PDF20180401HS.pdf.
406. Samet, J.M., et al., *Lung cancer in never smokers: clinical epidemiology and environmental risk factors.* Clin Cancer Res, 2009. **15**(18): p. 5626-5645.
407. Brown Johnson, C.G., J.L. Brodsky, and J.K. Cataldo, *Lung Cancer Stigma, Anxiety, Depression, and Quality of Life.* J Psychosoc Oncol., 2014. **32**(1): p. 59-73.
408. Cataldo, J.K. and J.L. Brodsky, *Lung Cancer Stigma, Anxiety, Depression and Symptom Severity.* Oncology, 2013. **85**(1): p. 33-40.
409. Chapple, A., S. Ziebland, and A. McPherson, *Stigma, shame, and blame experienced by patients with lung cancer: qualitative study.* BMJ, 2004. **328**(7454): p. 1470.
410. Marlow, L.A.V., J. Waller, and J. Wardle, *Does lung cancer attract greater stigma than other cancer types?* Lung Cancer, 2015. **88**(1): p. 104-107.
411. Sriram, N., et al., *Attitudes and Stereotypes in Lung Cancer versus Breast Cancer.* PloS one, 2015. **10**(12): p. e0145715-e0145715.
412. Else-Quest, N.M., et al., *Perceived stigma, self-blame, and adjustment among lung, breast and prostate cancer patients.* Psychol Health., 2009. **24**(8): p. 949-964.
413. Bresnahan Mary, J., K. Silk, and J. Zhuang, *You did this to yourself! Stigma and blame in lung cancer.* J Appl Soc Psychol, 2013. **43**(S1): p. E132-E140.
414. Eldridge, L. *Lung Cancer - The bias, the stigma, the shame, the blame and guilt* 2013; Archived at: http://www.webcitation.org/753Q84ZZR]. Available from: https://www.lungcancerfoundation.org/2013/06/lung-cancer-the-bias-the-stigma-the-shame-blame-and-guilt/.
415. Koskan, A., et al., *Use and Taxonomy of Social Media in Cancer-Related Research: A Systematic Review.* Am J Public Health, 2014. **104**(7): p. e20-e37.
416. Gualtieri, L. and F.Y. Akhtar. *Cancer patient blogs: How patients, clinicians, and researchers learn from rich narratives of illness.* in *Information Technology Interfaces (ITI), Proceedings of the ITI 2013 35th International Conference on*. 2013. .
417. Crannell, W.C., et al., *A pattern-matched Twitter analysis of US cancer-patient sentiments.* J Surg Res., 2016. **206**(2): p. 536-542.
418. Sutton, J., et al., *Lung Cancer Messages on Twitter: Content Analysis and Evaluation.* J Am Coll Radiol. , 2017.
419. *Facebook - Lung Cancer Support Group*. Available from: https://www.facebook.com/groups/733948110039512/?fref=nf.
420. Macmillan.org.uk. *Macmillan.org.uk - Lung Cancer Discussion Forum*. 2018; Available from: https://community.macmillan.org.uk/cancer_types/lung-cancer/.
421. Feinberg, J. *Wordle*. 2014; Available from: http://www.wordle.net.
422. Symplur. *#LCSM Tweet Chat*. 2018; Archived at: http://www.webcitation.org/6yLzA79Sp]. Available from: https://www.symplur.com/healthcare-hashtags/lcsm/.
423. Symplur. *#LungCancer healthcare social media hashtag*. 2018; Archived at: http://www.webcitation.org/6yM0JE70P]. Available from: https://www.symplur.com/healthcare-hashtags/lungcancer/.
424. Symplur. *Healthcare Social Media Research*. 2018; Archived at: http://www.webcitation.org/6yM0Rkw6G]. Available from: https://www.symplur.com/healthcare-social-media-research/.

425. Macmillan.org.uk. *Online Community News Blog: 100,000 members!* 2016; Archived at: http://www.webcitation.org/6yM0bwCvb]. Available from: https://community.macmillan.org.uk/blogs/b/community_news/archive/2016/02/10/100-000-members.
426. McNaught, C. and P. Lam, *Using Wordle as a supplementary research tool.* The Qualitative Report, 2010. **15**: p. 630+.
427. Langford Catherine Penny, H., et al., *Social support: a conceptual analysis.* J Adv Nurs., 2008. **25**(1): p. 95-100.
428. Wills, T.A., *Social support and interpersonal relationships*, in *Prosocial behavior.* 1991, Sage Publications, Inc: Thousand Oaks, CA, US. p. 265-289.
429. Uchino, B., *Social Support and Physical Health: Understanding the Health Consequences of Relationships.* 2004: Yale University Press.
430. Bossetta, M., *The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election.* Journalism & Mass Communication Quarterly, 2018. **95**(2): p. 471-496.
431. Wise, K., B. Hamman, and K. Thorson, *Moderation, Response Rate, and Message Interactivity: Features of Online Communities and Their Effects on Intent to Participate.* Journal of Computer-Mediated Communication, 2006. **12**(1): p. 24-41.
432. Balatsoukas, P., et al., *The Role of Social Network Technologies in Online Health Promotion: A Narrative Review of Theoretical and Empirical Factors Influencing Intervention Effectiveness.* J Med Internet Res 2015. **17**(6): p. e141.
433. #LCSM. *#LCSM Chat Topic 10/5: Lung Cancer Advocacy: How we can help ourselves*. 2017; Archived at: http://www.webcitation.org/71ReAIRM7]. Available from: https://lcsmchat.com/2017/10/04/lcsm-chat-topic-105-lung-cancer-advocacy-how-we-can-help-ourselves/.
434. Taylor, J., M. Osborne, and C. Pagliari. *Understanding the social dynamics of Twitter, Facebook and Diabetes.co.uk and their value implications for patients and health researchers*. GLOBAL HEALTH 2015 : The Fourth International Conference on Global Health Challenges 2015; Available from: https://www.thinkmind.org/download.php?articleid=global_health_2015_5_10.
435. Wiesenhofer, H., M. Ebner, and I. Kamrat, *Is Twitter an Individual Mass Communication Medium?*, in *Society for Information Technology & Teacher Education International Conference 2010*, D. Gibson and B. Dodge, Editors. 2010, Association for the Advancement of Computing in Education (AACE): San Diego, CA, USA. p. 1712-1717.
436. Macskassy, S.A., *On the Study of Social Interactions in Twitter.* ICWSM, 2012.
437. Kaplan, B.H., J.C. Cassel, and S. Gore, *Social Support and Health.* Medical Care, 1977. **15**(5): p. 47-58.
438. Gualtieri, L. and F.Y. Akhtar, *Cancer patient blogs: How patients, clinicians, and researchers learn from rich narratives of illness.* ITI 2013 35th International Conference on Information Technology Interfaces, 2013: p. 3-8.
439. Emanuel, E.J. and L.L. Emanuel, *The promise of a good death.* Lancet (London, England), 1998. **351 Suppl 2**: p. SII21.
440. Murray, S.A., et al., *Archetypal trajectories of social, psychological, and spiritual wellbeing and distress in family care givers of patients with lung cancer: secondary analysis of serial qualitative interviews.* BMJ, 2010. **340**.
441. Murray, S.A., et al., *Patterns of Social, Psychological, and Spiritual Decline Toward the End of Life in Lung Cancer and Heart Failure.* J Pain Symptom Manage., 2007. **34**(4): p. 393-402.
442. Scott, A.M., et al., *Exploring the spiritual needs of people dying of lung cancer or heart failure: a prospective qualitative interview study of patients and their carers.* Palliat Med., 2004. **18**(1): p. 39-45.
443. Holmberg, K., J. Bastubacka, and M. Thelwall, *"@God please open your fridge!" Twitter messages to @God in content analysis: Hopes, humor, spirituality, and profanities.* Journal of Religion, Media and Digital Culture, 2016. **5**(2): p. 339-355.
444. Lundby, K., *Patterns of belonging in online/offline interfaces of religion* Information, Communication & Society, 2011. **14**(8): p. 1219-1235.
445. Campbell, H.A., *Understanding the Relationship between Religion Online and Offline in a Networked Society*. Journal of the American Academy of Religion, 2012. **80**(1): p. 64-93.
446. Bruns, A. and J.E. Burgess, *The use of Twitter hashtags in the formation of ad hoc publics.* Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference, 2011.

447. Taylor, J. and C. Pagliari, *#Deathbedlive: the end-of-life trajectory, reflected in a cancer patient's tweets.* BMC Palliat Care. , 2018. **17**: p. 17.
448. Ko, H.-C., L.-L. Wang, and Y.-T. Xu, *Understanding the different types of social support offered by audience to A-list diary-like and informative bloggers.* Cyberpsychol Behav Soc Netw. , 2013. **16**(3): p. 194-199.
449. Boudewyns, V., et al., *Stigma's Effect on Social Interaction and Social Media Activity.* J Health Commun. , 2015. **20**(11): p. 1337-1345.
450. Occhipinti, S., et al., *Lung Cancer Stigma across the Social Network: Patient and Caregiver Perspectives.* J Thorac Oncol., 2018. **13**(10): p. 1443-1453.
451. Buckels, E.E., P.D. Trapnell, and D.L. Paulhus, *Trolls just want to have fun.* Personality and Individual Differences, 2014. **67**: p. 97-102.
452. Statista. *On which of the following topics have you seen trolling behavior on the internet?* 2018; Archived at: http://www.webcitation.org/6yLzPpeX8]. Available from: https://www.statista.com/statistics/380051/topics-witness-trolling-behavior-internet/.
453. Hall, A.K., et al., *The digital health divide: evaluating online health information access and use among older adults.* Health Educ Behav., 2015. **42**(2): p. 202-209.
454. Society, A.C. *Key Statistics for Lung Cancer*. 2018; Archived at: http://www.webcitation.org/6yLzl7IyB]. Available from: https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html.
455. Confessore, N., et al. *The Follower Factory - Everyone wants to be popular online. Some even pay for it. Inside social media's black market.* The New York Times 2018; Available from: https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html.
456. Ryan, G.W. and H.R. Bernard, *Techniques to Identify Themes.* Field Methods, 2003. **15**(1): p. 85-109.
457. Carter-Harris, L., *ES07.04 Leveraging Social Media to Change the Public Conversation on Lung Cancer Stigma.* Journal of Thoracic Oncology, 2018. **13**(10): p. S202.
458. Oscar Munoz-Garcia and C. Navarro, *Comparing user generated content published in different social media sources.* LREC Conference, 2012.
459. Taylor, J. and C. Pagliari, *Mining social media data: How are research sponsors and researchers addressing the ethical challenges?* Research Ethics, 2017: p. 1747016117738559.
460. Glaser, B.G., *Time for dying*, ed. A.L. Strauss. 1968, Chicago: Chicago : Aldine.
461. Strauss, L.A., *CHRONIC ILLNESS AND THE QUALITY OF LIFE*. 1976. p. 82-82.
462. Murray, S.A., et al., *Illness trajectories and palliative care.* Murray , S A , Kendall , M , Boyd , K & Sheikh , A 2005 , ' Illness trajectories and palliative care ' BMJ , vol 330 , no. 7498 , pp. 1007-11 . , 10.1136/bmj.330.7498.1007, 2005.
463. Reed, E. and J. Corner, *Defining the illness trajectory of metastatic breast cancer.* BMJ Supportive & Palliative Care, 2015. **5**(4): p. 358.
464. Tang, M.H., D.J. Castle, and P.F.M. Choong, *Identifying the Prevalence, Trajectory, and Determinants of Psychological Distress in Extremity Sarcoma.* Sarcoma, 2015. **2015**.
465. Segal, S., N. Katherine, and P. Laune, *Psychological adaptation, coping and distress in adult onset soft tissue sarcomas.* Electronic sarcoma update newsletter, 2010. **7**(4).
466. Taubert, M., et al., *Palliative social media.* BMJ Supportive & Palliative Care, 2014. **4**(1): p. 13.
467. Bernardo, T.M., et al., *Scoping review on search queries and social media for disease surveillance: a chronology of innovation.* Journal of medical Internet research, 2013. **15**(7): p. e147.
468. Charles-Smith, L.E., et al., *Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review.(Report).* 2015. **10**(10).
469. Wikipedia. *Kate Granger*. [cited 2017 2017-01-08]; Archived at: http://www.webcitation.org/6nMXfhAls]. Available from: https://en.wikipedia.org/wiki/Kate_Granger.
470. Granger, K. *@GrangerKate*. [cited 2016 2016-09-11]; Archived at: http://www.webcitation.org/6nMXuFRXm]. Available from: https://twitter.com/grangerkate?lang=en.
471. Alliance, S. *End-of-Life Issues*. 2017; Archived at: http://www.webcitation.org/6pmrtPmNX]. Available from: http://sarcomaalliance.org/what-you-need-to-know/end-of-life-issues/.
472. Varis, P. *Digital Ethnography*. Tilburg Papers in Culture Studies 2014 [cited 2016; Archived at: http://www.webcitation.org/6nMYwbv44]. Available from: https://www.tilburguniversity.edu/upload/c428e18c-935f-4d12-8afb-652e19899a30_TPCS_104_Varis.pdf.

473. Hsieh, H.-F. and S.E. Shannon, *Three approaches to qualitative content analysis.* Qualitative health research, 2005. **15**(9): p. 1277.
474. Hymes, D.H., *Ethnography, linguistics, narrative inequality : toward an understanding of voice*. 1996, London: London : Taylor & Francis.
475. Barclay, S., et al., *Living in uncertain times: trajectories to death in residential care homes.* British Journal of General Practice, 2014. **64**(626): p. e576.
476. Murray, S.A., et al., *Patterns of Social, Psychological, and Spiritual Decline Toward the End of Life in Lung Cancer and Heart Failure.* Journal of Pain and Symptom Management, 2007. **34**(4): p. 393-402.
477. Southall, D., *The patient's use of metaphor within a palliative care setting: Theory, function and efficacy. A narrative literature review.* Palliative Medicine, 2012. **27**(4): p. 304-313.
478. Gajewski, M. *May I take your metaphor? – how we talk about cancer*. 2015; Archived at: http://www.webcitation.org/6uv8Rajsg]. Available from: http://scienceblog.cancerresearchuk.org/2015/09/28/may-i-take-your-metaphor-how-we-talk-about-cancer/.
479. Mindruta, R. *Marketing: Top 15 Free Social Media Monitoring Tools*. 2017; Archived at: http://www.webcitation.org/6uv9CUepd]. Available from: https://www.brandwatch.com/blog/top-10-free-social-media-monitoring-tools/.
480. Cronin, G. *Living and dying on social media*. 2015; Archived at: http://www.webcitation.org/6pmrtPmNX]. Available from: https://www.cilip.org.uk/blog/living-dying-social-media.
481. Calvard, J. *Death and Grief Online: The Opportuntiies and Challenes of Incorporating Digital Legacies into Palliative Care in Hospice Settings*. 2016; Archived at: http://www.webcitation.org/6pn1kO28z]. Available from: http://www.stcolumbashospice.org.uk/wp-content/uploads/2016/11/Digital-legacy-presentation.pdf.
482. Shaw, D.M., J.V. Gross, and T.C. Erren. *Why You Should Donate Your Medical Data When You Die: Organs are not hte only item of value from the deceased*. The Conversation 2017; Archived at: http://www.webcitation.org/6pn21BKUB]. Available from: https://www.scientificamerican.com/article/why-you-should-donate-your-medical-data-when-you-die/.
483. Meier, E.A., et al., *Defining a Good Death (Successful Dying): Literature Review and a Call for Research and Public Dialogue.* Am J Geriatr Psychiatry, 2016. **24**(4): p. 261-71.
484. Taylor, J. and C. Pagliari, *Mining social media data: How are research sponsors and researchers addressing the ethical challenges?* Research Ethics, 2017. **14**(2): p. 1-39.
485. Leidner, J.L. and V. A Plachouras. *Ethical by Design: Ethics Best Practices for Natural Language Processing*. in *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. 2017. Valencia, Spain.
486. Gibney, E. *The battle for ethical AI at the world's biggest machine-learning conference*. 2020 [cited 2020; Archived at: http://archive.is/umClA]. Available from: https://www.nature.com/articles/d41586-020-00160-y.
487. IDF. *IDF Diabetes Atlas*. 2019 [cited 2020; Archived at: http://archive.is/1H1eG]. Available from: https://www.diabetesatlas.org/en/sections/worldwide-toll-of-diabetes.html.
488. Taylor, J. and C. Pagliari, *The social dynamics of lung cancer talk on Twitter, Facebook and Macmillan.org.uk.* npj Digital Medicine, 2019. **2**(1): p. 51.
489. Falisi, A.L., et al., *Social media for breast cancer survivors: a literature review.* Journal of Cancer Survivorship, 2017. **11**(6): p. 808-821.
490. Jed R. Brubaker, F.K.-S., Lee Taber, and Gillian R. Hayes. *Grief-Stricken in a Crowd: The Language of Bereavement and Distress in Social Media*. in *Sixth International AAAI Conference on Weblogs and Social Media*. 2012.
491. Lingel, J., *The Digital Remains: Social Media and Practices of Online Grief.* The Information Society, 2013. **29**(3): p. 190-195.