Applications of Agent-based Models in Tobacco Control

Adarsh Prabhakaran

Artificial Intelligence and its Applications Institute
School of Informatics
University of Edinburgh

This dissertation is submitted for the degree of
Doctor of Philosophy

October 2024

To my grandparents …

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Adarsh Prabhakaran
October 2024

Acknowledgements

I am deeply grateful to my supervisors, Dr. Valerio Restocchi and Dr. Ben Goddard, for their unwavering support during my four-year journey. Despite the challenges often faced by a supervisor's first PhD student, my experience under Valerio's guidance has been nothing short of exceptional. His office door was always open; I could simply drop by to say 'hi' and discuss the progress of our work. Similarly, Ben has been consistently accessible and incredibly detail-oriented, greatly enhancing the quality and rigour of my research. Both of them ensured my progress and helped me step back to see the bigger picture whenever I found myself lost in a rabbit hole.

I would also like to express my thanks to Dr. Jennifer Badham for serving as external examiner and to Prof. Jacques Fleuriot for his invaluable feedback during the annual reviews and as the internal examiner for my thesis. Both examiners provided insightful feedback that significantly enhanced the final version of this thesis.

I offer my special gratitude to our collaborators, Prof. Jamie Pearce, Dr. Roberto Valiente Borox, and Dr. Garth Reid. The discussions in our meetings often exceeded what traditional academic literature could offer.

I appreciate the PHASE network and AIAI for providing financial grants and conference attendance support, respectively.

I extend my heartfelt thanks to all the members of AIML who always welcomed me into their research group meetings and provided invaluable feedback on my work. Forming friendships beyond the confines of the office has been invaluable during my

PhD. Guillermo, Jorge, Zonglin, James, Resul, Ondrej, and Ginny, my colleagues turned dear friends, have made this journey even more special.

A PhD is a unique journey where the line between work and personal life often blurs. Having friends who understand both your professional and personal sides has been very helpful. Ameer, Arushi, Octave, Nikita, and Tarini, I'm grateful to have met you. Our coffee breaks, dessert runs, and dinners have been among my most cherished experiences over the last four years.

Despite the challenges of moving to a new country right before a global pandemic, Edinburgh has been incredibly welcoming. The friends and experiences this city has offered have been invaluable. I've been particularly fortunate to meet a group of wonderful people - Aparna, Aswin, Anirudh, Mugdha, Swetha, Devisree, Junaina, Kartik, Kavita, Neelakshi, Smrithi, Sooraj, and Vinothini. Thank you all for making this journey so much more fulfilling and making Edinburgh feel like home.

Over the years, I've been a part of several circles in Edinburgh, but it's Anisha, Awanish, Dayal, Kanak, Meera, and Nikhil who have truly left an imprint on me. You all have been more than just friends; you've been intellectual companions, sparking intense discussions on society and politics that have both comforted me and ignited my curiosity. Thank you for enriching my time here.

I also wish to thank my closest friends who have been a constant source of support throughout the PhD and before. Priyasha, for always being the best support system. Varun and Awani for keeping me grounded. Vishnu and Dubey for always being there when I needed them. Physical distance has never diminished our friendship; rather, it has only highlighted how integral you all are to my life.

Finally, I extend my deepest gratitude to my parents and siblings, whose boundless support and understanding have been my backbone throughout this PhD journey. A special mention must go to my brothers Arun and Adithya, who have been a constant source of motivation. I must also give heartfelt thanks to my grandparents, who instilled in me values that have been my moral compass. Though the past few years have been marked by the loss of dearly beloved ones, their teachings and love continue to light my way. Thank you for shaping me into who I am today.

Abstract

Tobacco use remains one of the leading public health threats worldwide. Despite the implementation of various policies, the tobacco epidemic still persists, particularly in deprived communities. The tobacco use environment is a complex system, with multiple factors interacting with each other. These factors include not only the interactions among individuals that contribute to the spread of smoking behaviour but also the influence of socio-economic and spatial environments on individual smoking habits. This thesis employs a complex systems perspective to enhance our understanding of the tobacco use environment. Utilising agent-based models (ABM), we aim to capture the various factors contributing to different smoking processes and behaviours. Through these models, we investigate the spread of smoking behaviour, evaluate the impact of network topology on these behaviours, develop novel network-based interventions to curb smoking, explore the socio-economic and spatial complexities affecting smoking initiation, and assess the diverse long-term impacts of tobacco regulations.

The first part of the thesis focuses on the spread of tobacco use through interactions between individuals, using a network-based ABM. We demonstrate that network topology and specifically network degree, is a significant factor in the spread of smoking behaviour. Furthermore, we suggest alternative synthetic networks for instances when network data is unavailable. Utilising the model, we can devise network interventions aimed at leveraging the contagious nature of quitting behaviour.

In the subsequent chapter, we shift our focus to the development of a simple, cost-effective network intervention to enhance the effectiveness of smoking cessation clinics. Our proposed approach is based on the ‘friendship paradox’, which suggests one’s friends are likely to have more friends than oneself. Our findings indicate that this friendship paradox based strategy outperforms existing methods employed in smoking cessation programmes, both in lowering overall smoking rates and in reducing the disparity between smoking levels in deprived and non-deprived areas.

The third chapter introduces a novel spatial ABM that captures the complexities of smoking initiation. This framework considers not just interactions between individuals but also how socio-economic and spatial environments influence smoking behaviour. We simulate a realistic neighbourhood in Scotland, replicating daily commutes, social interactions, and smoking patterns among agents. Our model draws upon extensive data from Scotland to craft a comprehensive socio-economic landscape, incorporating elements such as population density, retailer distribution, cigarette pricing, age demographics, school and workplace densities, income brackets, and existing smoking rates. With a focus on the 15 to 24 age group, a critical demographic for smoking initiation, this chapter provides an adaptable simulation platform for assessing the long-term consequences of various types of regulatory strategies.

In the final chapter, the spatial ABM developed is employed to assess the impact of four distinct tobacco control policies: raising cigarette prices, enforcing a ban on tobacco sales around schools, increasing the legal age of sales to 21, and a strategy that combines all three. To assess the impact of these policies, we examine multiple tobacco indicators such as prevalence, tobacco consumption, tobacco sales, and financial spending on tobacco. We also compare the effects of these policies across different socio-economic neighbourhoods. We show that a multifaceted approach not only enhances the impact of individual policies but also effectively reduces the widening of socio-economic inequalities.

Collectively, this thesis showcases the potential of ABMs as a dynamic testbed for comprehending, simulating, and shaping tobacco control strategies. The findings presented herein contribute to the advancement of evidence-based tobacco control policies by offering a multifaceted perspective on tobacco use dynamics and the complexities of the regulatory environment.

Contents

Chapter 1

Introduction

The global tobacco use epidemic is one of the most pressing and silent public health challenges as well as a global health priority (World Health Organization et al., 2023). Every year, more than eight million people worldwide die due to tobacco-related diseases, ranging from respiratory issues to cancers (GBD 2019 Tobacco Collaborators and others, 2021; Reitsma et al., 2017). Although many countries have implemented strategies like public media campaigns (Vallone et al., 2009), restrictions on advertisement for cigarettes (Gilpin and Pierce, 1997; Rose et al., 2013), health warnings, cessation support (Brown et al., 2014b) and control on access to tobacco products (Millett et al., 2011; Schneider et al., 2011) to curb tobacco use (Flor et al., 2021), the epidemic persists, with higher prevalence in deprived communities. This puts a strain on health systems as well as has significant socio-economic implications. The socio-economic consequences of tobacco use exacerbate already existing health-equity gap, further marginalising communities for whom healthcare is less accessible due to socio-economic status (Jha et al., 2006; Stringhini et al., 2010)

The tobacco environment is a complex with multiple elements including tobacco users, tobacco retailers, and even non-tobacco users (Wallace et al., 2015). A wide range of factors, including age demographics (O'Loughlin et al., 2014), financial means (Cui et al., 2019), socio-economic status (Hitchman et al., 2014) and the simultaneous interactions between these factors, can influence this system. These factors play an important role in different processes in tobacco use behaviour including social influence, social normalisation, availability and access to tobacco products (Liu et al., 2017; Shortt et al., 2016). Tobacco use has been termed a “wicked problem” due to its complex nature, the limited/lack of knowledge about different aspects of the issue as well as the socio-economic burden it imposes (Dorfman and Wallack, 1993; Young et al., 2012). It is not surprising that this has led to huge efforts from different research communities to characterise the system and study it further.

Computational models have recently emerged as one of the vital tools for understanding this complex system and for evaluating impact of tobacco control interventions (Feirman et al., 2017). Various approaches have been taken in this direction. However, most of these do not account for the contagious nature of tobacco use behaviour. Research has shown that tobacco use and smoking behaviour can be seen as a social contagion, where the behaviour spreads through interactions between tobacco users and non-tobacco users (Block et al., 2020; Christakis and Fowler, 2008).

In this thesis, we leverage a complex systems approach and use existing empirical understanding of tobacco social contagion to address the following research themes:

1. Can social contagion models serve as platforms for developing and testing new policies aimed at curbing tobacco use?
2. Can models incorporating social contagion enhance our understanding of the tobacco use landscape and its varied impacts?

1.1 Motivation

Smoking prevalence has witnessed a significant global decrease over the past few decades. Worldwide, smoking prevalence among adults declined from over 40.0% in the 1980s to 22.8% in 2020 (Ng et al., 2014; World Health Organization and others, 2019, 2021). This positive trend can be attributed to the collaborative efforts of many countries through the United Nations Framework Convention on Tobacco Control (UN FCTC) (World Health Organization and others, 2004), which has been instrumental in developing and implementing effective tobacco control policies like restriction of tobacco advertisement (Gilpin and Pierce, 1997; Rose et al., 2013), controlling the access to tobacco (Millett et al., 2011; Schneider et al., 2011) and increasing cessation support (Brown et al., 2014b). In the case of Scotland and the UK, specific policies have played a crucial role in contributing to the reduction in smoking initiation and prevalence in young individuals. The Smoking Health and Social Care Act, which prohibits smoking in enclosed public spaces (Scottish Executive, 2005), and the Children and Young Persons (Sale of Tobacco, etc.) Act, which increased the minimum legal age to purchase tobacco from 16 to 18 (Department of Health, 2007), have been significant drivers of these positive outcomes (Anyanwu et al., 2020; Millett et al., 2011). These policies reflect a comprehensive strategy to combat smoking initiation and prevalence, ultimately leading to a healthier population.

However, despite this progress, tobacco use remains one of the most significant preventable causes of illness and death on a global scale (GBD 2019 Tobacco Collaborators and others, 2021). Persistent socio-economic disparities are evident, with higher smoking prevalence in deprived areas compared to non-deprived ones (ASH Scotland, 2014). Furthermore, although there has been an increase in quitting attempts through national smoking cessation services, the declining number of individuals accessing these clinics points to the ongoing challenges that need to be addressed (ASH Scotland, 2014). This highlights the importance of continued efforts to tackle tobacco use and its associated health disparities.

1.2 Challenges

1. **The Complexity of Tobacco Use**

   The tobacco use landscape involves multiple stakeholders, from individuals (both users and non-users) and products to retailers and policymakers, creating a complex environment (Wallace et al., 2015). The interactions and inter-dependencies of these diverse players complicate the tobacco regulatory environment. This environment epitomises a complex system, which can be defined as a set of interconnected elements that produce behaviour and properties not easily predicted from the individual parts alone (Arthur, 1999; Ladyman et al., 2013; Weaver, 1948). In such a system, even though the individual components have their own properties, due to the complex interactions new behaviours and patterns emerge.

In the context of tobacco use, this complexity has significant implications for policy and intervention. Aiming a policy at a single component of this system can, due to unforeseen interactions, lead to outcomes that might not be as impactful as intended, or even worse, could produce unintended detrimental effects. This can transpire in various forms, such as a surge in smuggling (HM Revenue & Customs , 2018) and illegal sales (Chartered Trading Standards Institute 2015, 2015), a transition of users to cheaper and potentially more harmful cigarettes due to higher tobacco prices (Gilmore et al., 2015; Shortt et al., 2021), heightened financial strain on already deprived individuals (Blakely and Gartner, 2019; Guillaumier et al., 2015; Hoek and Smith, 2016), or even an amplification of stigmatisation associated with smoking (Riley et al., 2017). In essence, the challenge is not just about regulating tobacco but understanding and navigating the maze of complexities inherent to its use and control.

2. **The Social Contagion of Smoking Behaviour**

   Tobacco use, particularly smoking, is not just a standalone individual decision but is deeply intertwined with broader social influences. The environments in which individuals find themselves in and the people they interact with heavily influence their smoking behaviours (Christakis and Fowler, 2008). A large body of research demonstrates the significant influence of social ties on smoking behaviours. From initiation to cessation, an individual's decision to smoke or quit is often a reflection of the behaviours observed within their close social circles (Blok et al., 2017; Ennett et al., 2008; Mercken et al., 2009). Close family and friend relationships exert dual influences on smoking patterns. On one hand, non-smokers are more likely to start smoking if someone close to them smokes. On the other hand, smokers are more encouraged to quit if someone in their circle does the same (Christakis and Fowler, 2008). Conversely, those who remain insulated from smokers in their close relationships tend to stay non-smokers, viewing non-smoking as the default norm. Such patterns of behaviours adopted based on interpersonal relationships transform smoking from a mere personal choice to a social contagion. A nuanced understanding of this social contagion can be instrumental in crafting effective anti-smoking policies. Leveraging the contagious nature of behaviour presents a challenge that, if addressed, could lead to the formulation of policies designed to amplify the spread of cessation behaviours while minimising smoking initiations.

3. **Unavailability of Network Data**

   The contagious nature of a smoking behaviour through interpersonal relationships and social ties, make the social network of individuals very important (Burgess-Hull et al., 2018; Väänänen et al., 2008). Despite the critical role that network information plays in understanding smoking behaviours, a significant data gap remains. At larger population levels, comprehensive data on the social network aspects of smoking and other similar social contagions remain limited. To effectively leverage the contagious nature of smoking for positive interventions, there is an urgent need for broader research efforts. Detailed studies focusing on the social network dimensions of these contagions can provide invaluable insights and pave the way for more targeted and impactful interventions. The scarcity of detailed network data thus presents an urgent challenge that must be addressed to enable more targeted and impactful policy interventions.

4. **The Role of Socio-economic Factors in Smoking Behaviour**

   Socio-economic status, similar to an individual's social ties, can be a significant predictor for smoking behaviours. Empirical evidence points out that, individuals in more deprived areas have much higher prevalence of smoking than those in less deprived areas (of National Statistics, 2014). This implies that individuals in most-deprived areas are exposed to smokers more (Hitchman et al., 2014), which can lead to more normalisation of smoking, and also high effects of social contagion. Further, deprived areas have been associated with higher density of tobacco retailers (Shortt et al., 2015). A higher density of tobacco retailers implies more availability of tobacco products as well as higher accessibility.

This in turn again affects the exposure to tobacco products and users (Caryl et al., 2020). The stark socio-economic disparities in smoking prevalence and accessibility to tobacco products present another pressing challenge, underscoring the need for targeted policies that account for these inequities.

5. **Implications for Policy and Intervention**

   For policymakers, grasping the intricate interplay between social ties, socio-economic status, and the broader tobacco landscape is essential. This complexity underscores the need for multifaceted interventions that consider not just the individual smoker, but their environment and influences as well. This raises a formidable challenge for policymakers: crafting interventions that are nuanced enough to address the complex, interconnected factors that contribute to tobacco use, rather than treating each factor in isolation.

1.3 Contributions

We aim to address the above challenges in our this thesis. Our efforts have resulted in the following contributions:

- **Tobacco social contagion model**: We address Challenges 1, 2 and 3 in the first contribution by developing an agent-based network model to simulate the dynamics of smoking initiation, cessation, and relapse. Using the model on various network structures, we identify that average degree is the most important network property for smoking dynamics. Additionally, we suggest alternate synthetic networks when real-world network data is not available. This contribution has been peer-reviewed and published in the journal Applied Network Science (Prabhakaran et al., 2023) and has been presented as an extended abstract at the Complex Networks conference (Prabhakaran et al., 2022).

- **Network-based intervention strategies**: Collaborating with Public Health Scotland, we utilised the tobacco social contagion model to develop an innovative network based strategy aimed at enhancing the efficiency of smoking cessation programs, addressing challenges 2, 3 and 5. Simulating our simple heuristic succeeds in reducing smoking prevalence across diverse socio-economic backgrounds while also being highly cost-effective . This strategy is currently under consideration for implementation in Scotland.

- **Spatial heterogeneous agents framework**: To tackle Challenges 1 and 4, we developed a comprehensive agent-based framework that models the interactions leading to smoking initiation, through factors like peer influence, parental guidance, and addiction. This model captures interactions across different age groups and integrates nuanced behaviours like mobility and purchasing habits with retailers. Using extensive data from diverse socio-economic areas from Scotland, our framework serves as a simulation framework, where policy makers can gauge 'what-if' scenarios of their strategies across various smoking and socio-economic indicators.
- **Evaluation of tobacco control policies**: In collaboration with Public Health Scotland, we used the spatial heterogeneous agents framework to evaluate the impact of four distinct policies on multiple tobacco use indicators, thereby addressing Challenges 1, 4 and 5. These policies include raising the price of tobacco, enforcing a ban on the sale of cigarettes near schools, raising the minimum legal age for tobacco purchases to 21, and a combination of all three. Our findings suggest that employing a multifaceted approach targeting different factors simultaneously is more effective at reducing tobacco use than implementing individual policies in isolation. Moreover, this approach promotes greater equity. The significance of this research was highlighted through its presentation to officials from the Government of Scotland, Public Health Scotland, ASH Scotland, and various other stakeholders at a workshop, emphasising its real-world implications for tobacco control policies in Scotland. Unlike the network-based intervention strategy, which focuses on leveraging social dynamics for cessation programs, this contribution provides a broader evaluation of population-level policy impacts, offering valuable insights for policymakers in designing comprehensive tobacco control strategies.

1.4 Thesis outline

This thesis is structured as follows:

1. Chapter 1: Introduction to the thesis.
2. Chapter 2: Development of an agent-based model to study tobacco social contagion on networks.
3. Chapter 3: Application of the tobacco social contagion model from Chapter 2 to devise novel network-based interventions.
4. Chapter 4: Development of a heterogeneous spatial agents model. This model simulates the behaviours of different age groups, taking into account factors such as susceptibility to peer pressure, mobility, and purchasing habits.
5. Chapter 5: Utilising the spatial model developed in Chapter 4, we test four distinct policies. The impact of these policies on tobacco use and socio-economic indicators is analysed.
6. Chapter 6: A summary of the research findings, discussions on their implications, and suggestions for future work.

Chapter 2

Network-based Social Contagion Model

2.1 Introduction

Smoking is one of the leading preventable causes of death, disability and disease across the world(Office for National Statistics, 2019; US Department of Health and Human Services and others, 2014, 2020) and it is one of the most significant avoidable hazard factors for cancer (Banks et al., 2015) and respiratory diseases (Ferkol and Schraufnagel, 2014). Not only is smoking a global health burden, but it is also an economic burden, which significantly outweighs the economic benefits from tobacco production and sales (Drope et al., 2018).

Acknowledging the need for active efforts on tobacco control, in 2005 182 countries ratified the first international public health treaty, the WHO Framework Convention on Tobacco Control (FCTC) (World Health Organization and others, 2004). This involved cigarette taxation, smoke-free zones, public media campaigns (Vallone et al., 2009), restrictions on advertisement for cigarettes (Gilpin and Pierce, 1997; Rose et al., 2013), health warnings, cessation support (Brown et al., 2014b) and control on access to tobacco products (Millett et al., 2011; Schneider et al., 2011). Effective implementation of such tobacco control policies has decreased deaths, extended both the lifespan and life expectancy of the population (Holford et al., 2014; van Meijgaard and Fielding, 2012), and is also associated with a predicted decrease in healthcare expenditure (Lightwood and Glantz, 2016). However, despite the success of these tobacco control policies, smoking is still prevalent with the world's economies still spend more than one trillion USD per year on smoking-related health expenditures and loss of productivity (Acharya et al., 2016; Goodchild et al., 2018).

An important factor that is commonly overlooked by tobacco-control policy making is the long-term effect of social contagion on both smoking levels and socio-economic inequalities, especially given that the latter can be unexpectedly exacerbated by policies that only optimise short-term effects (Caryl et al., 2021).

In fact, despite the large body of evidence that suggests that tobacco initiation and cessation largely depend on social ties (Blok et al., 2017; Christakis and Fowler, 2008; Ennett et al., 2008; Go et al., 2010; Mercken et al., 2009), there is currently no model that fully captures the complexity of social contagion in smoking dynamics.

Consequently, such a model is crucially needed to develop policies that not only accurately take into account long-term effects, but also exploit social contagion to enhance tobacco control.

This is not to say models of social contagion for smoking dynamics do not exist. In fact, many models have been proposed in which smoking is compared to a disease spreading in a population due to its contagion-like behaviour (Sharomi and Gumel, 2008; Zaman, 2011a,b; Zaman et al., 2017). This similarity allows the use of tools from epidemiology to model the propagation of smoking behaviour, leading most models to be variations of compartmental models such as SIR and SIER that use ordinary differential equations (ODEs) to describe the smoking dynamics (Castillo-Garsow et al., 1997; Sharomi and Gumel, 2008). However, epidemiology has successfully moved on to the more flexible agent-based models (ABM) which can incorporate individual-level heterogeneity in behaviours, spatial dynamics and complex social networks, (Aleta et al., 2022; Hunter et al., 2018; Thurner et al., 2020), but tobacco control has not. Therefore, existing models of social contagion for tobacco control cannot accurately reproduce the empirically observed complexity of this phenomenon, due to the following reasons.

First, most of these models do not account for the topology of the social ties. When modelling any type of social contagion, it is known that the social interaction between individuals plays an important role (Hodas and Lerman, 2014; Shin, 2022). Therefore such models can only be accurate if they account for the underlying social network of the population. Even though compartmental models try to incorporate social interactions between groups of individuals, there is no network structure involved. To reduce the complexity and analytical tractability, these models ignore the structure of the social ties and instead assume a homogeneous well-mixed population, which means that any individual can infect others in the system (Anderson et al., 1992; Kermack and McKendrick, 1927, 1932, 1933). This means that vital information from the actual social network is not taken into consideration (Moore and Newman, 2000).

Second, when these models do consider network structure, they arbitrarily fix the structure. The models of smoking behaviour which consider the structure of these social ties are agent-based models (Chao et al., 2015; Schaefer et al., 2012, 2013). In these cases, the topology is usually arbitrarily fixed as a scale-free network or on small scale school-network. A scale-free network, characterised by a power-law distribution of connections where a few nodes have many connections and most of them have few (Barabási and Albert, 1999), is commonly used to approximate large-scale social networks. Since real-world social contact networks of adults can be very different from synthetic and school networks, there is a need for careful characterisation of the smoking behaviour of these models on different network topologies.

Finally, even though empirical research shows that smoking initiation and cessation largely depend on social ties (Christakis and Fowler, 2008), not all of these interactions have been considered for modelling the spread of smoking. Although one of the first theoretical studies which modelled smoking cessation advocated the use of interactions between smokers and quitters (Castillo-Garsow et al., 1997), interactions which lead to smoking cessation and relapse are still not used (Schaefer et al., 2013). Consequently, in these models smoking cessation and relapse are usually determined only by spontaneous terms, which causes an underestimation of social contagion effects (Sharomi and Gumel, 2008; Zaman, 2011a,b; Zaman et al., 2017). As previously discussed, this underestimation can lead to unintended consequences of tobacco control policies such as an increase in socio-economic inequalities.

A closely related field to the spread of smoking behaviour is opinion dynamics, wherein mathematical and computational models are used to study the spread of opinions in a population by considering social influence. Like opinions, smoking behaviour can be influenced by the attitudes and behaviours of others, such as peers, family members, and by external influences like media (Colaiori and Castellano, 2015; Mueller and Tan, 2018). However, unlike opinions, smoking is a health hazard. This makes the spread of the smoking behaviour also similar to that of an infectious disease where instead of the infection, smoking behaviour is the contagion. Therefore, modelling smoking behaviour can be seen through a hybrid lens of epidemiology and opinion dynamics. Such an approach will help us combine the opinion dynamics perspective which considers the social and psychological factors that influence the adoption and maintenance of smoking behaviour (such as peer pressure, familial attitude towards towards smoking, etc.) and the epidemiological perspective capturing the mechanisms contributing to the spread of smoking-related health hazards.

Over the years, the effect of social networks on individual and population behaviour have been studied in both opinion dynamics and epidemiology (Rahmandad and Sterman, 2008). Multiple approaches have been designed to study the spread of diseases and opinion according to the situation and amount of information available. More recently, these models have been used to study a variety of social contagions including obesity (Hill et al., 2010b), emotions (Hill et al., 2010a), alcoholism (Lee et al., 2010; Sharma and Samanta, 2015), substance abuse (White and Comiskey, 2007), behaviour change (Badham et al., 2021), and information spreading (Zhou et al., 2020), to name a few. Due to the similarity of the spread of smoking behaviour to other social contagions, we can use insights from these fields to develop models.

To address these issues with existing smoking dynamics models, we develop an agent-based model (ABM). ABMs are a class of computational techniques which rely on dynamical interactions between autonomous agents to understand the emerging properties of a complex system due to these local interactions. We use ABMs for the following reasons:

First, unlike the ODE models, ABMs can easily be extended to multiple theoretical and real-world networks. This versatile nature allows ABMs to be applied to different population structures and more realistic models of the system. Therefore, the effect of network structure on the smoking dynamics can easily be studied by changing the underlying network topology. Consequently, we can characterise the dynamics of smoking behaviour in multiple different network topologies.

Second, interactions between agents can easily be incorporated into ABMs without significantly increasing the complexity of the model. Conversely, when multiple interactions are considered in an ODE model, the model becomes too complex to be solved analytically. By constructing ODE models which are not solvable (or ones in which the solutions are too complex and lengthy), it becomes difficult to validate and analyse the nature of the solutions analytically. Instead, such ODE systems must be solved using numerical methods to approximate the solutions.

In our ABM for smoking, we include three state change processes: smoking initiation, cessation and relapse. Each of these processes can occur due to both interactions and spontaneously. This accounts for the multiple possible interactions which can lead to a change in smoking behaviour.

Finally, ABMs are flexible enough to become effective test-beds for developing new policies. One of the main applications of studying contagion (both infectious diseases and social contagion) is to develop strategies to contain them. Effective strategies and interventions may prevent smoking initiation, motivate smokers to quit, and stop former smokers from relapsing. Due to the socially contagious nature of smoking, we can potentially use network-based strategies developed from studies on infectious diseases and other social contagions.

First, we develop an agent-based model, which considers multiple possible interactions between heterogeneous agent groups, along with spontaneous terms accounting for external influences, to study the spread of smoking behaviour. Our model can be used to develop innovative intervention strategies and policies for tobacco control. Furthermore, we show the robustness of our ABMs by comparing the dynamics against a traditional ODE model and as expected, our results suggest that our ABM on a fully connected network and the equivalent ODE model provide the same results. Additionally, we show that ABMs on fully connected networks and ODE models should not be used to model smoking behaviour as they replicate the real-world data with poor accuracy compared to the other networks.

Next, we explore the effect of the underlying network topology on smoking dynamics. We find that the underlying network structure affects smoking dynamics considerably. However, synthetic networks with the same average degree reproduced the historic data and showed similar characteristics as that of the real-world networks. Specifically, we show that Lancichinetti-Fortunato-Radicchi benchmark networks and random networks can be used to develop intervention strategies when complete information on the underlying network topology of a local population is not available.

2.2 Methods

This section describes the model structure, data, and the networks used, along with the modelling choices involved in developing the ABM for smoking behaviour.

We highlighted the need for a tobacco control model to incorporate both spontaneous and interaction terms, as well as to consider appropriate network topologies. To achieve this, we use a synthetic population of n (please note the use of small case n, since N will be used to denote never-smokers) agents in an undirected and unweighted network G. To show the effects of the network topology, we make the agents interact on six different networks (described later in the section) and compare the observed dynamics.

2.2.1 Description of the agent-based model

In our model, each agent can be in one of the following smoking states: never-smoker (N), smoker (S) or quitter (Q). An agent is a never-smoker if they have never smoked before, while an agent who smokes any tobacco product daily, or occasionally, falls into the smoker state. Finally, if a smoker quits smoking even temporarily, they are labelled a quitter. We initiate the agents into each of the above states randomly.

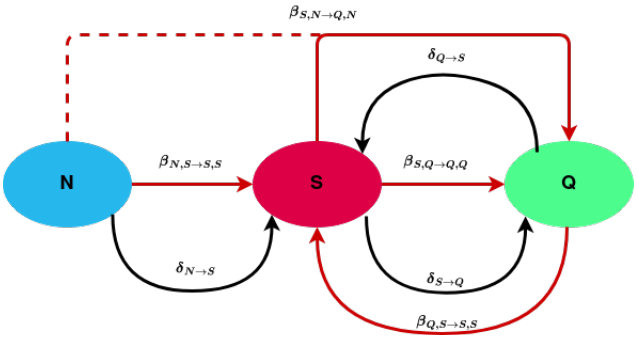


Figure 2.1: The figure shows the schematic representation of the state change processes involved in the ABM. The interaction parameters are represented by the red arrows, while the black arrows show the spontaneous terms in the schematic. All three state-change processes are shown in the figure. First, an N-agent can initiate smoking spontaneously ($\delta_{N\to S}$) or due to the interaction with an S-agent ($\beta_{N,S\to S,S}$). Similarly, an S-agent can quit spontaneously ($\delta_{S\to Q}$) or due to interaction with other non-smoker agents (Q-agent: $\beta_{S,Q\to Q,Q}$ or N-agent: $\beta_{S,N\to Q,N}$). Like the other processes, Q-agents relapse into smoking spontaneously ($\delta_{Q\to S}$) or due to interaction ($\beta_{Q,S\to S,S}$) with an S-agent.

To make models more realistic over long time periods, epidemiological models usually include vital dynamics. Traditionally, this is done by including constant mortality and birth rates in the equations. In traditional network-based models, a constant birth rate and mortality rate can lead to older agents having a higher number of social contacts (for simpler models), thus increasing their influence on other agents. However, in reality, the number of social connections does not increase with age but instead peaks in the mid-twenties and then decreases with age (Bhattacharya et al., 2016). However, tackling these problems would significantly increase the complexity of the model since the model will have to include age-dependent mortality and birth rates, and the network generation process will have to be adapted to ensure that the network retains its properties while adding and removing agents. Hence, we do not include vital dynamics in the model and run our experiments in time frames where the vital dynamics can be ignored. We calibrate and validate our model in time periods of less than 30 years to minimise the effect of vital dynamics.

State change processes

We incorporate three main processes into the model: smoking initiation, smoking cessation and relapse into smoking. The three state change processes involve interaction-based state changes as well as spontaneous state transitions. We assume that each exposure to an agent with a different smoking status is independent of the previous exposure. We then use a binomial approximation to compound the effect of multiple independent interactions on one agent simultaneously to calculate the probability of state change. The binomial approximation provides a tractable way to calculate interaction probabilities in a system with multiple interactions. Similar approximations have been used in modelling the spread of diseases and behaviours in populations (Keeling and Rohani, 2011).

In the following subsection, we provide a detailed description and derivation of the approximation.

Binomial approximation for independent exposure

Suppose agent i with state Z has n_i neighbours, of which n_i^Y are in state Y. Let X be the event in which the agent i interacts with a Y neighbour, this can have two outcomes, success = (agent i becomes a Y agent), or failure = (agent i stays a X agent). Let b be the probability of successful interaction of agent i with the Y agent and k be the number of successes, then the probability of k successes is:

$$P(k) = \binom{n_i^Y}{k} b^k (1-b)^{(n_i^Y - k)} \tag{2.1}$$

Probability of state change is then:

$$\begin{aligned} P &= P(k > 0) && (2.2) \\ &= 1 - P(k = 0) && (2.3) \\ &= 1 - (1-b)^{n_i^Y} && (2.4) \end{aligned}$$

This value only focuses on the total number of Y agents and does not take into account the total number of neighbours the agent i has. (Blok et al., 2017) shows that more than the number of individuals in a particular state (smokers or quitters), it is the proportion of them in your relationship circle that has an impact on an individuals smoking behaviour. In order to incorporate the size of the network neighbours, we scale it with the fraction of Y neighbours to the total number of neighbours. Therefore the probability for smoking initiation due to interaction is

$$\beta_{N,S\to S,S} = \frac{n_i^S}{n_i}(1-(1-b)^{n_i^S}) \tag{2.5}$$

Similarly,the probability of smoker agent i quitting due to interaction with quitters is:

$$\beta_{S,Q\to Q,Q} = \frac{n_i^Q}{n_i}(1-(1-b)^{n_i^Q}) \tag{2.6}$$

Probability of smoker agent i quitting due to interaction with never-smokers is:

$$\beta_{S,N\to Q,N} = \frac{n_i^N}{n_i}(1-(1-b)^{n_i^N}) \tag{2.7}$$

Probability of quitter agent i relapsing due to interaction with smokers is:

$$\beta_{Q,S\to S,S} = \frac{n_i^S}{n_i}(1-(1-b)^{n_i^S}) \tag{2.8}$$

Smoking initiation: First and foremost, we define the transition of a never-smoker into a smoker as smoking initiation. In the model, an N-agent can initiate smoking in two ways. First, through a random probability $\delta_{N\to S}$ depicting various external influences like advertisements, movies, and the presence of tobacco shops influencing an N-agent to pick up smoking. Second, through interactions with other S-agents in its network neighbourhood with a probability $\beta_{N,S\to S,S}$. We use the binomial approximation mentioned before to calculate the expression in (2.9), which gives the probability of smoking initiation due to interaction.

$$P_{N,S\to S,S} = \frac{n_S}{n}(1-(1-\beta_{N,S\to S,S})^{n_S}) \tag{2.9}$$

Smoking cessation: Following smoking initiation, we define the process of an S-agent quitting smoking as smoking cessation. Similar to smoking initiation, smoking cessation can also happen in two ways. First, due to various external influences like mass-media campaigns, mandatory warning labels on cigarette boxes, and higher taxes. This external influence is incorporated into the model through the spontaneous term $\delta_{S\to Q}$. Second, due to interactions with non-smokers. However, both Q-agents as well as N-agents fall under the non-smoker category. Therefore unlike the other state change processes, interactions with both the other states can lead to smoking cessation. $\beta_{S,N\to Q,N}$ represents the probability of cessation of an S-agent due to N-agents in its network neighbourhood. At the same time, the probability of cessation due to other Q-agents in its network neighbourhood is given by $\beta_{S,Q\to Q,Q}$. Subsequently, the probability of smoking cessation due to interaction of an S-agent with multiple Q-agents is given by (2.10) while (2.11) gives the same probability but due to interaction with multiple N-agents.

$$P_{S,Q\to Q,Q} = \frac{n_Q}{n}(1-(1-\beta_{S,Q\to Q,Q})^{n_Q}) \tag{2.10}$$

$$P_{S,N\to Q,N} = \frac{n_N}{n}(1-(1-\beta_{S,N\to Q,N})^{n_N}) \tag{2.11}$$

Smoking relapse: Finally, we define picking up smoking after a period of abstinence as a smoking relapse. Smoking relapse is similar to smoking initiation, except that the Q-agent gets influenced instead of an N-agent. Similar to the other two cases, smoking relapse can happen in two ways. $\delta_{Q\to S}$ represents the probability of a Q-agent relapse into smoking due to external influence. Additionally, $\beta_{Q,S\to S,S}$ represents the probability of Q-agent relapsing into smoking due to interaction with its immediate network-neighbour S-agents. Here, like the smoking initiation, only interactions with other S-agents can cause a Q-agent to relapse into smoking. The probability of Q-agent relapsing due to its interaction with multiple S-agents is shown in (2.12).

$$P_{Q,S\to S,S} = \frac{n_S}{n}(1-(1-\beta_{Q,S\to S,S})^{n_S}) \tag{2.12}$$

Experiment settings

We run simulations with a total population of $n = 1000$ agents as this population size allows for multiple simulations and parameter explorations within reasonable computational time constraints. These agents are connected with each other based on the pre-defined network structure. To understand how this pre-defined network structure affects smoking behaviour, we vary the network structure and study the smoking dynamics observed in each network. This process involves comparing the model on each network with the empirically observed data. Along with this comparison, we also identify the combination of parameter values that best fit observed data and how this combination changes with the underlying network. In addition, we also compare the results of the ABM on different networks with the ODE analogue (described in supplementary section ODE model) of the ABM.

In the ABM, each time step corresponds to a year in the real world as most smoking related data are reported yearly or bi-yearly making the calibration of the model easier. At every timestep, agents follow a three-step procedure sequentially to avoid cascading agent states. First, each agent identifies a potential new state it can transition to based on the agent's state at that time step. Next, each agent calculates the probability of transitioning into the identified new state. Finally, each agent simultaneously transitions into the new state based on the calculated probability. The sequential procedure above removes the chance of cascading agent states in a single step. That is, an N-agent can never change into an S-agent and then a Q-agent in the same time step.

Due to the stochastic nature of ABMs, we iterate each simulation multiple times. We iterate each simulation ten times during the parameter sweeps for calibrating the model as full parameter sweeps involve testing numerous combinations, making higher iterations time-consuming and computationally expensive. However, to validate the model, we iterate the best-fit combinations 1000 times to increase the likelihood of observing and accounting for low-probability events.

For every simulation, a new network is generated, that is, each simulation has a different realisation of the network structure. We then initiate $s_0\%$ and $q_0\%$ (based on the first data-point in the empirical data) of the total 1000 agents randomly as S-agents and Q-agents, respectively, on the generated network.

The model was built using the modular framework Mesa in Python (Kazil et al., 2020). We chose Mesa for its flexibility in customising model components, efficient handling of large-scale simulations and parallel processing, and seamless integration with Python's data analysis libraries.

2.2.2 Networks

We have run simulations of our ABM on six different network topologies: fully-connected, scale-free (Barabási and Albert, 1999), random (Erdős et al., 1960), small world (Watts and Strogatz, 1998), Lancichinetti-Fortunato-Radicchi benchmark (Lancichinetti et al., 2008) and a real-world network from the Framingham heart study (FHS) data (Hill et al., 2010b). These networks were chosen due to their unique properties or ubiquitous nature in literature. The scale-free network, random network and small-world networks are standard network topologies in network sciences that are used for spreading phenomenons; hence, we test our model on these networks too. Details of each of the networks used are mentioned below.

1. Fully-connected (FC) network: The fully-connected network assumes that every agent is connected to every other agent in the system. This is to replicate the mean-field or perfect-mixing approximation seen in ODE models. In smoking modelling, this represents an idealised scenario where everyone influences everyone else, functioning as a maximum threshold for behaviour spread. However, real-world networks are sparse and seldom fully-connected. Therefore, we generate and explore other network topologies.

2. Erdős–Rényi (ER) network: Similar to the FC network, every node in a random network tries to form an edge with every other node, but with a probability p_{er}. The situation when $p_{\text{er}} = 1$ corresponds to a FC network. We use the Erdős–Rényi (ER) model (Erdős et al., 1960) to generate random networks for our experiments. In studies of epidemiology, ER networks are often used as a neutral benchmark to isolate the effects of network structure on dynamics. We adopt this approach for smoking modelling, using the ER network as a baseline to compare against more structured network topologies.

3. Barabási–Albert (BA) network: These are networks where the degree distribution follows a power law. Many real-world networks have been reported to follow the power-law distribution (Albert et al., 1999; Gamermann et al., 2019). To model this we use the Barabási–Albert (BA) network model (Barabási and Albert, 1999). In the context of smoking behaviour, BA networks model the presence of highly influential individuals who can impact smoking initiation or cessation rates in their large social circles, thus capturing the enormous influence of some individuals in real-world social networks.

4. Watts–Strogatz (SW) network: This is a network which is highly clustered with small average shortest paths. These networks are known for local cliques and random long-ranged connections. We use the Watts–Strogatz model to generate the network (Watts and Strogatz, 1998). SW networks capture how smoking behaviours might spread quickly through seemingly distant social groups, reflecting real-world phenomena like workplace or school interactions where local groups are connected by occasional long-range ties.

5. Lancichinetti-Fortunato-Radicchi (LFR) benchmark network: The LFR network encompasses properties of a real-world network like a heterogeneous distribution of degrees and size of communities (Lancichinetti et al., 2008). We use an LFR network due to its unique property of communities embedded into it during the network generations process. Thus, LFR networks allow us to test how

community structures within social networks might influence the spread of smoking behaviours.

6. Framingham Heart Study (FHS) network: We have also used a real-world network based on the Framingham heart study (FHS) (Dawber, 2013) data along with the synthetic networks above. The FHS was a a longitudinal cohort-based study aimed explicitly at studying cardiovascular diseases and identifying the associated factors. However, due to the wide range of documented associated factors, it has become a one-of-a-kind data set on which even detailed network analysis has been carried out. We used a configuration model to generate synthetic networks with the same degree distribution observed in the FHS (Hill et al., 2010b). Unlike the SF and LFR networks, the configuration model allows an arbitrary distribution of degrees and is therefore not restricted to the power-law distribution (Newman, 2003).

Each of the synthetic networks (ER, BA, SW, and LFR) mentioned above has multiple parameters associated with its generation. We chose these parameters involved in the network generation process such that their average degree is close to the empirically observed values from social network studies, particularly those related to smoking behaviour. A study conducting extensive network analysis using the FHS data, which models a a similar social contagion reports average degrees ranging between 2.8 and 5.3 over different years (Hill et al., 2010b). Other studies on the composition of social networks in other countries have observed an average degree of 4.45 (Chao et al., 2015). However, the network involved in the spread of smoking behaviour mainly consists of close family members and friends (Christakis and Fowler, 2008). Therefore, the number of individuals potentially influencing such behaviour will be less than the average degree of a standard social network. For this reason, we choose an average degree of $\langle k \rangle = 3$, based on exam 6 of the FHS, for the generation of networks in our model. In cases where algorithmic restrictions prevent us from using $\langle k \rangle = 3$ (such as with the SW model), we use the closest possible value, which is $\langle k \rangle = 4$.

Style	Mean Degree	Degree Variance	Clustering Coefficient	Mean Geodesic
BA	3.9920	34.4999	0.0332	3.9555
ER	3.1140	3.1430	0.0018	6.0847
FHS	3.0220	6.4035	0.0013	5.3656
LFR	3.9020	301.1884	0.1397	3.1209
SW	4.0000	1.0360	0.1612	6.1052

Table 2.1: Network properties for the different network types tested using the model.

The parameter values for each network and the average expected degree are shown in Table 2.1. For each network type, we report four key metrics: Mean Degree ($< k >$), Degree Variance (σ_k^2), Clustering Coefficient (C), and Mean Geodesic ($\bar{l}$).

These metrics provide insights into the structural characteristics of each network type. Notably, we see that the BA and LFR networks show higher σ_k^2, indicating more heterogeneous connectivity. The SW network exhibits the highest C, reflecting its local clustering property. The ER and SW networks have larger $\bar{l}$, suggesting longer average path lengths. The FHS network shows the lowest $\bar{k}$ and C among all network types.

2.2.3 Data

To calibrate and validate the model, we use publicly available smoker and quitter prevalence data from the US and UK. The UK data-set (Office for National Statistics, 2019) has normalised smoker population and quit ratio (defined as the proportion of smokers who have quit smoking) from 1974 to 2019. Bi-yearly data points are available until 2000, and yearly data from then on. We estimate the quitter population from the quit ratio and use it to calibrate the model.

In the case of the US, we used the data available in the official Surgeon General's report on tobacco (Office of United States Public Health Service and others, 2020). This document reports the prevalence of smokers (male and female separately) and quitters (again, male and female separately) between the years 1965 and 2015 (data points every five years).

Additionally, we impose the values of two parameters ($\beta_{N,S\to S,S}$ and $\beta_{S,Q\to Q,Q}$) of the model by estimating them from empirical research (Christakis and Fowler, 2008). We describe the steps taken to estimate the parameters below.

Parameter estimation from Christakis et al. The paper by Christakis et al. (Christakis and Fowler, 2008) quantifies the relative probability of an individual picking up smoking and quitting smoking due to interaction with a close social tie using the Framingham heart study (FHS) data set. We use the values from the paper to calculate the interaction parameters $\beta_{N,S\to S,S}$ and $\beta_{S,Q\to Q,Q}$.

The FHS study is a longitudinal observational cohort study initiated to study various risk factors involved with cardiovascular diseases. The Christakis paper looks at 7 exam waves and calculates the relative probabilities using the method of logistic regression. The paper gives the relative increase in probability of smoking initiation and cessation due to interaction. According to our definitions in the model, this implies that the interaction probabilities for smoking initiation and smoking cessation are related to their respective spontaneous terms.

Since our model does not differentiate between social relationships, we take a weighted average of the relative probability considering the total number of members falling in each of the different social ties mentioned in the paper. We consider the average probability from the seventh wave (the final one) for our calculations.
From the paper, we have calculated these values as

$$\begin{aligned} g &= 0.35214 \\ b &= 0.40719 \end{aligned} \tag{2.13}$$

Where g and b are the weighted average of the probabilities of smoking cessation and initiation respectively (averaged considering different social tie).
As mentioned in the main paper, our model used a time unit of 1 year. However, the Christakis paper calculates the probabilities based on a regression between seven unequally split examinations (unequal in the time space). The detailed derivation of the steps used to calculate the yearly probabilities is given below.

Calculating yearly probabilities from probability over multiple years Suppose the initial population of smokers is S_0 and the probability of quitting in a year is p. Let, Q_t be the number of people who will quit in year t and S_t be the number of people remaining in smoker compartment in year t.

Then,

$$\begin{aligned} S_1 &= (1-p)S_0 \\ S_2 &= (1-p)^2 S_0 \\ &\vdots \\ S_t &= (1-p)^t S_0 \end{aligned} \tag{2.14}$$

Therefore the number of quitters at every time step becomes,

$$\begin{aligned} Q_1 &= pS_0 \\ Q_2 &= pS_1 = p(1-p)S_0 \\ &\vdots \\ Q_t &= pS_{t-1} = p(1-p)^{t-1}S_0 \end{aligned} \tag{2.15}$$

The total number of quitters due to this state change becomes

$$\begin{aligned} Q_{total} &= \sum_{i=1}^{t} Q_i \\ &= pS_o(1 + p(1-p) + p(1-p)^2 + \ldots + (1-p)^{t-1}) \\ &= pS_0 \sum_{i=1}^{t} (1-p)^{t-1} \\ &= S_0(1-(1-p)^t) \end{aligned} \tag{2.16}$$

The paper reports the probability of quitting over t years which is equivalent to:

$$\begin{aligned} g &= \frac{Q_{total}}{S_o} \\ &= 1-(1-p)^t \end{aligned} \tag{2.17}$$

Therefore the probability of quitting per year is,

$$p = 1 - (1-g)^{1/t} \tag{2.18}$$

This can be generalised for both smoking initiation and smoking cessation.

By applying 2.18 to both the interaction term and spontaneous term, we can calculate the interaction probability per year as

$$\begin{aligned} \beta_{S,Q\to Q,Q} &= 1 - \left[1 - g(1-(1-\delta_{S\to Q})^t\right]^{1/t} \\ \beta_{N,S\to S,S} &= 1 - \left[1 - b(1-(1-\delta_{N\to S})^t\right]^{1/t} \end{aligned} \tag{2.19}$$

We assume that, each person got examined in the same order, and take the average of the time between the starting year of each examination to calculate the time between two exams. We estimate the value of $t = 4.5$.

2.2.4 Calibration and validation

To reproduce the smoking prevalence trends observed in the UK and the US, we calibrate and validate the model using a four-step process. To limit the effects of not including vital dynamics, we calibrate and validate the model on time frames of 25-30 years. Then, we split the time-stamped data into calibration and validation segments in UK and US scenarios. We use smoking prevalence data from 1974 to 2002 (16 data points) in the calibration segment and the remaining in the validation segment in the UK data. Similarly, in the US, we use the data from 1965 - 1990 to calibrate the model and the remaining to validate it.

Step 1: Coarse-grained calibration: To identify the combinations of parameter values which best mimic the empirical prevalence data, we run a coarse-grained parameter sweep on all the uncertain parameters of the ABM. In this case, this parameter sweep was carried out for each parameter using ten logarithmically split values between 0 and 1, allowing us to efficiently cover a wide range of magnitudes from very small values to those approaching 1. Further, we iterated each parameter combination ten times to reduce the effects of randomness. We then use these simulation results to identify the range of parameters best fitting the calibration data (top 100 best-fit parameter combinations). Since we have the population sizes time-series data, we use the sum of the Mean Square Error (MSE) of both the S and Q trends to identify the best fitting parameters.

Step 2: Sensitivity analysis: After identifying the set of parameters that minimise the MSE, we perform a sensitivity analysis on these values to determine the relative influence of each model parameter on the smoking dynamics. This involves individually varying each parameter across its range while holding other parameters constant and evaluating their impacts on the simulated smoker and quitter prevalence curves over time. Despite varying levels of sensitivity, we determined that all parameters contribute to capturing the complex dynamics of smoking behaviour for all network structures tested. Therefore, we retained all original parameters in the final model, as each plays a role in accurately representing the behaviour under various conditions. Details on the sensitivity analysis results, including plots of varying each parameter and discussion of their relative impacts on prevalence, are available in Supplementary Section Sensitivity analysis.

Step 3: Fine-grained calibration: To improve the estimated parameters, we recalibrate the model through a finer grained parameter sweep on each parameter. In this case, this parameter sweep was carried out for each parameter for five equally split values between the range of values identified in step 1 for improving parameter resolution while keeping the computational costs low. Just as in step 1, we iterate each of the simulations ten times. We use the range of values identified for each parameter from the coarse-grained calibration to run the fine-grained parameter sweep.

Step 4: Validation: We validate the calibrated model by comparing the simulated results with the validation data for both US and the UK. Since we used the sum of MSE of the smokers and quitters for calibrating the model, we also use the same sum of MSEs for validation. Along with the MSEs, we also use the unique crossover point of the historical trends of the smoker and quitters populations to improve the validation process.

2.3 Results

In this section, we study the ABM for smoking and show its characteristics. First, we compare the ABM with all possible interactions on different networks. Through this comparison, we demonstrate the importance of networks for modelling the spread of smoking behaviour. We further demonstrate the ease with which networks can be incorporated into ABMs and therefore advocate using them for modelling such a spreading phenomenon. Next, we compare the ABM with an ODE analogue to demonstrate the equivalence of the ABM on an FC network and a traditional ODE model. Finally, we calibrate and validate the model on empirical data observed in the US and the UK. Through the calibration and validation process, we emphasise the need to incorporate networks into such models, which can potentially be used to develop policies. Next, we show that the real-world network (FHS) replicates the empirical data observed in the US and the UK. In addition, we show that in practical situations, when complete information on the actual underlying network is not available, synthetic networks with similar average degrees can be used to develop models. On the other hand, our results suggest that ABMs on FC networks, and by extension ODE models, may not adequately capture the complex dynamics of smoking and similar behavioural contagions. We examine the evolution of total prevalence of smokers (S) and quitters (Q) for each simulation setup. To compare and quantify the temporal dynamics of the populations, we calculate the sum of MSE of the S and Q curves. Since the S and Q curves cross each other in both the UK and US data, we study the unique crossover time-point for the ABM and compare it to the one from empirical data.

2.3.1 Population dynamics

Figure 2.2 shows the population dynamics observed from simulations of the ABM for smoking (on all six network topologies) and an ODE analogue (described in

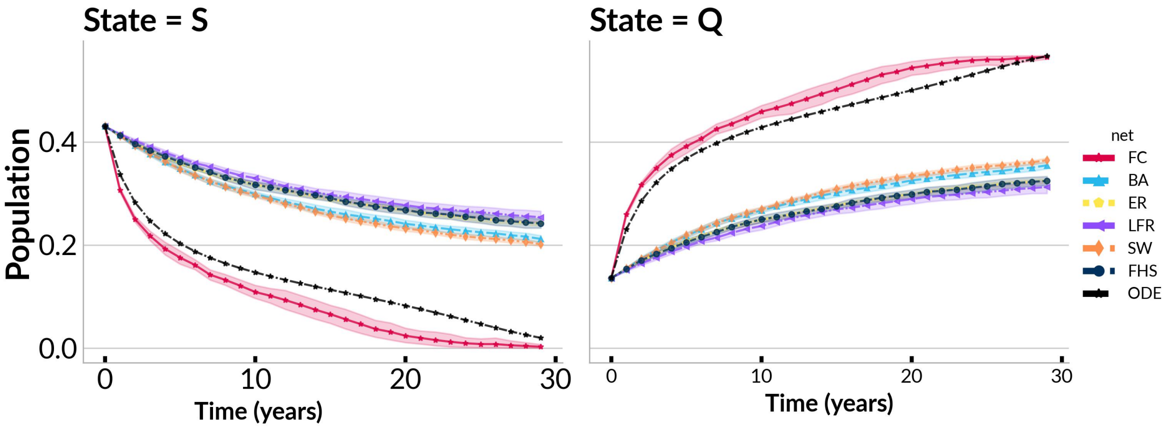


Figure 2.2: Dynamics of the S and Q populations from the ODE model and ABM on networks (FC, BA, ER, FHS, LFR and SW networks). We used the best-fit parameters from the coarse-grained parameter sweep of the FHS network $(\beta_{Q,S\to S,S}, \beta_{S,N\to Q,N}, \delta_{N\to S}, \delta_{S\to Q}, \delta_{Q\to S} = 0.01334, 0.05623, 1e-05, 4e-05, 1e-05)$ for these simulations. The black dotted lines in both the plots represent the ODE model results for the respective population for the same parameters mentioned above. The ODE model and ABM on FC show similar temporal dynamics, while the dynamics change drastically when the network structure moves away from FC.

Supplementary section ODE model) with the same parameter values. We run the ABM and ODE models to simulate a period of 30 years. This time frame is similar to what we used to calibrate the models on US and UK data. We use this observed population dynamics to compare the ABM between different networks and the ABM on these networks with the ODE model.

Our results suggest that the network structure affects the population dynamics of smokers and quitters for the same experimental conditions. Specifically, when the network structure is changed from FC to any other network, the dynamics observed change drastically. In addition, when the average degree of the networks (other than in the FC network) is kept at similar values (close to $\langle K \rangle = 3$), we observe that the deviation in the observed population dynamics is minor between different networks. This deviation being minor suggests that the average degree is vital in the dynamics of smoking behaviour as expected from the model definition.

The ABM on a fully-connected network follows the same qualitative trajectory as the ODE model. Statistical equivalence of a differential equation model, which uses a mean-field approximation to an ABM on a fully-connected network, has been shown before, and our model is consistent with this result (Rahmandad and Sterman, 2008). This suggests that while simple traditional ODE model can be used to model smoking behaviour and other similar social contagions, it may have limitations when applied to populations with network structures different from a fully-connected topology.The effectiveness of ODE models in such cases would depend on how significantly the actual network structure influences the dynamics of smoking behaviour.

2.3.2 Validation

When the parameters that best captured the historical trend for the ABM on each network were observed, the variation in the MSE values was small. Therefore to make the parameter selection process more robust, we chose 100 combinations that gave the minimum MSE values instead of choosing only the minimum MSE parameter combination. Additionally, we imposed a condition that each independent parameter in this combination had to fall between the first and third quartile of its values observed in the minimum 100 (the values which fall inside the box in Figure 2.5). We then sample this new set of parameters 1000 times to validate the model.

To validate the calibrated model, we compare the evolution of simulated population sizes with the historical data. Since we calibrate the model using the MSE values of S and Q curves, we also use the same for validation. For validation, we use periods of the empirical data, which were not used to calibrate the model (validation data, periods: 2003-2019 for the UK and 1995-2015 for the USA).

Case 1: UK

Figure 2.3 shows the distribution of MSE values, crossover points and the population dynamics of 1000 iterations of the simulations using the best-fit parameters for each network in the UK. We compare the MSE distributions of each network with each other in terms of their central tendencies and their spread. The FC network showed the largest difference in MSE distribution compared to all other networks, with notably higher mean and median values. This aligns with the poor performance of FC observed during validation. The ER network followed by the FHS network had the smallest

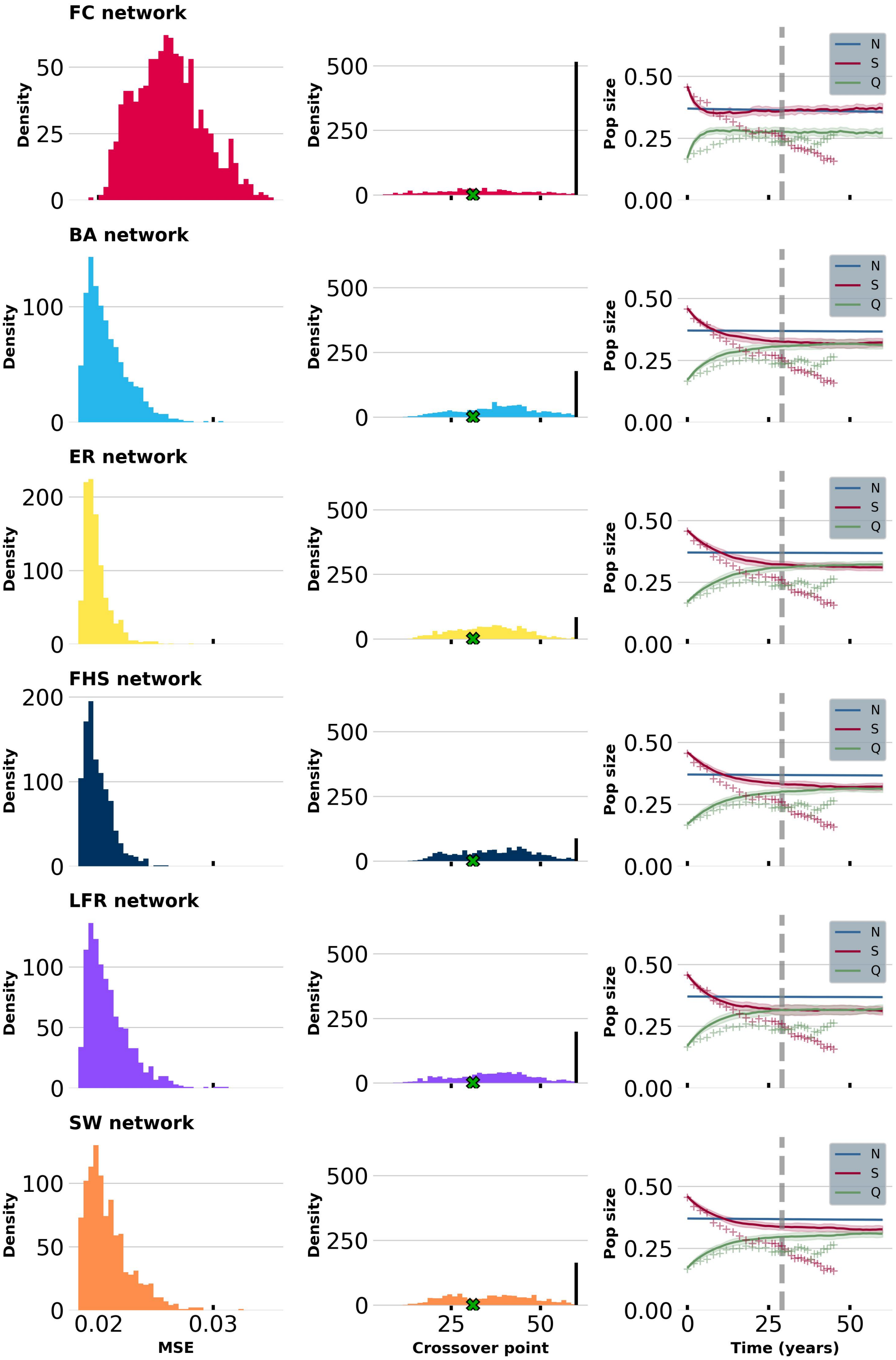


Figure 2.3: The figure shows simulated characteristics and population plots from 1000 runs of the ABM for the best-fit parameters on each network for the UK.

Figure 2.3: Each column represents a different characteristic: the first column shows the MSE value (sum of S and Q) of the ABM with the validation data, the second column shows the crossover point between the smoker and quitter populations, and the third column shows the mean population value with a 95% confidence interval (CI) around the curve, obtained from 1000 simulation runs. Each row represents a different network type (FC, BA, ER, FHS, LFR, SW). The green crosses on the second column on the x-axis represent the actual crossover point in the empirical data. The black bars in the second column show the number of times the S and Q curves do not cross each other. The third column shows the mean simulated population curves over time for smokers (S) and quitters (Q) from 1000 runs, with the shaded areas indicating the 95% CI. The lines with the marker $+$ indicate the actual historical prevalence data. The grey dotted line, dividing the plot indicates the time-step till which the model was calibrated. Among the six networks, we see that ER (MSE mean = 0.01984, SD = 0.00069) and the FHS (MSE mean = 0.01998, SD = 0.00071) network reproduces the data most accurately. The BA (MSE mean = 0.02076, SD = 0.00104), SW (MSE mean = 0.02081, SD = 0.00135) and LFR (MSE mean = 0.02082, SD = 0.00114) are very similar to each other in terms how good they replicate the data. While the ABM on FC network (MSE mean = 0.02617, SD = 0.00226) provides the worst fit for the validation data.

MSE values with a much lower standard deviation. While the BA network's MSE distribution showed considerable overlap with those of the LFR and SW networks.

The distributions of crossover points in Figures 2.3 and 2.4 (row 2) show the time steps at which the simulated smoker and quitter curves intersect over the 1000 runs (if there is a crossover) for each network topology. The width of these distributions illustrates the variability in when the crossover occurs between different iterations of the model parameters. The range of crossover points observed in the distribution always contains the actual crossover point observed in the historic trend. However, the number of times the crossover happens are very different when the underlying network is changed. The S and Q curves cross over 485, 822, 915, 912, 802, 836 times for FC, BA, ER, FHS, LFR and SW respectively. The FC network both gives a higher variability in the crossover point distributions and a lower number of successful crossovers, while the ER gives the highest number of successful crossovers. The third row in Figure 2.3 shows the mean simulated population curves over time for smokers and quitters, along with the 95% confidence interval (CI) bounds. The CI illustrates the range of dynamics observed across the 1000 iterations of the best-fit model parameters for each network. Additionally, the lines with the '+' marker depict the actual historical smoking prevalence data from the UK for comparison to the model results. From the difference in MSE values and the variability in the crossover-points we can conclude that the network structure does matter when modelling the smoking behaviour. The real-world FHS network and the ER network replicate the historic trend observed very well and also has the highest number of successful crossovers.

For the spread of smoking behaviour, the influence of other individuals is only substantial when they are a close family member or a close friend (Christakis and Fowler, 2008). This limits the average degree of the required network for the spread of smoking behaviour. This network will have a much lower average degree than a standard social network. In such situations, the degree distribution of the ER network can approximate real-world ones, such as that of the FHS network. This can be seen in our results, the ER and FHS network best replicate the validation data out of the six networks. The LFR and SW networks also generate low MSEs. This trend is seen in the number of successful crossovers as well. Following the MSE distributions, the FHS and ER also gives very high successful crossovers, while BA, LFR and SW give similar values of successful crossovers. Additionally, the ABM on FC networks gives the worst fit to the validation data and the least number of successful crossovers showing that the ABM on FC networks and, by extension, ODE models should not be used to model smoking and similar behavioural contagion. We can conclude that ABM on the ER network can be potentially used to model smoking behaviour when information on real-world network is not available.

Supplementary section A.4 gives a detailed analysis of the networks used for the simulations.

Case 2: US

Figure 2.4 shows the distribution of MSE values, crossover points and the population dynamics for the ABM over 1000 iterations using the best-fit parameters calibrated with the US data. As in the case of the UK, we compared the MSE distributions for each network in the US, focusing on their central tendencies and spread. The distribution of MSE values for the FHS network, followed by that of the LFR network, had a smaller spread and lower median and mean values. The SW, BA, and ER networks followed in order of increasing mean MSE values. As in the UK, the FC network performed worst, with the highest mean and median MSE values.

Even though the distribution of MSE values were significantly different, the average MSE values of the ABM on each of the networks were lower than that of the ABM on FC and very close to each other. The third row in Figure 2.4 shows the mean simulated population curves over time for smokers and quitters, along with the 95% confidence interval (CI) bounds. We find that the ABM calibrated on the FHS network provides the overall best fit to the empirical smoking data for both the US. However, the models using the LFR, BA, ER, and SW networks also match the historic trends. We find that the FHS network fits the real-world data the best in the case of the US, followed by the LFR network. It's worth noting that the LFR network, unlike the others, has a community structure embedded in its generation process. While this suggests that community structure might play a role in the spread of smoking behaviour. Given the limitations of our model and the relatively small differences in fit between network types, further research would be needed to confirm the importance of community structure in smoking dynamics.

Analysis of best-fit parameters

Figure 2.5 shows a box plot of values seen in the 100 best-fit parameters for each network on both US and UK data. When the 100 best parameter combinations that fit the data best were compared, the FC network consistently gave significantly different parameter combinations in both US and UK data sets. In both US and UK, at the 5% significance level, the values of at least 80% of the parameter values found from calibration on the FC network were significantly different from the other networks. Moreover, 100% of them were significantly different from the FHS network. On the other hand, all other networks return parameter values in which at least 20% of them are not significantly different from that of the FHS. The only exception is the SW network in the case of the US, where all parameters values returned were significantly different.

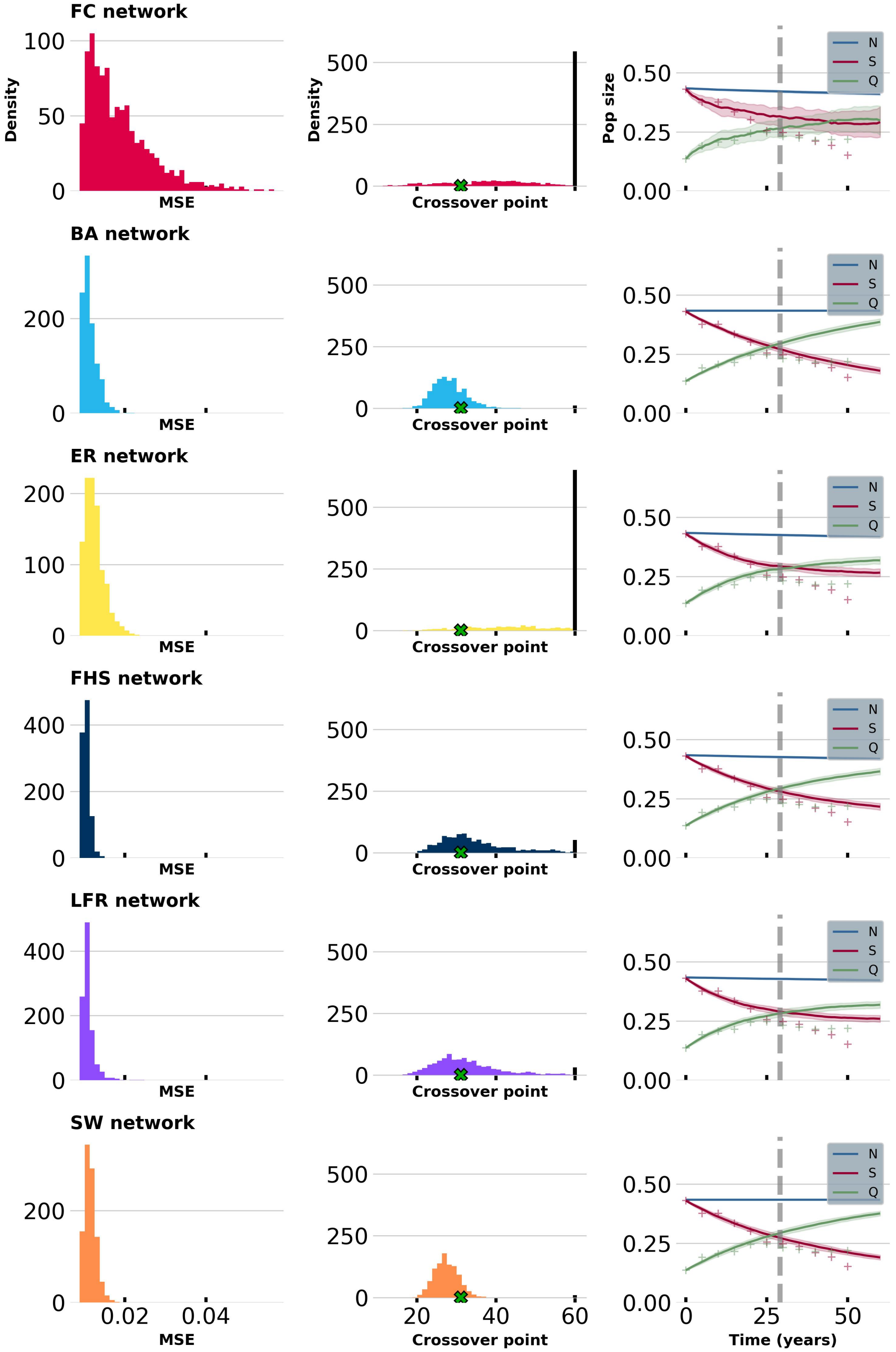


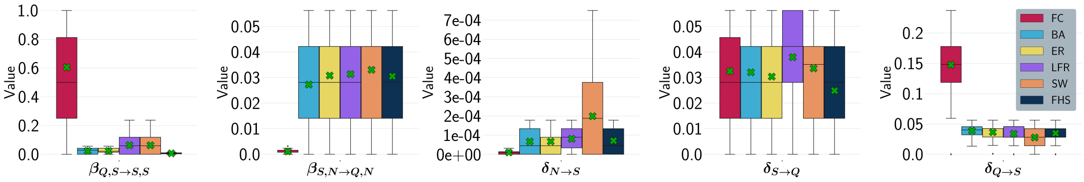
Figure 2.4: The figure shows simulated characteristics and population plots from 1000 runs of the ABM for the best-fit parameters on each network for the US.

Figure 2.4: Each column represents a different characteristic: the first column shows the MSE value (sum of S and Q) of ABM with the validation data, the second column shows the crossover point, and the third column shows the mean population plot with a 95% CI. Each row represents a different network type (FC, BA, ER, FHS, LFR, SW). The green crosses on the second column on the x-axis represent the actual crossover point in the empirical data. The black bars in the second column show the number of times the S and Q curves did not cross each other. The third column shows the mean simulated population curves over time for smokers (S) and quitters (Q) from 1000 runs, with the shaded areas indicating the 95% CI. The lines with the marker $+$ indicate the actual historical prevalence data. The grey dotted line, dividing the plot indicates the time-step till which the model was calibrated. Amongst the six networks, we see that the ABM on the FHS network reproduces the data most accurately (MSE mean = 0.01048, SD = 0.00053). The LFR (MSE mean = 0.01102, SD = 0.00091), BA (MSE mean = 0.0114, SD = 0.0013) and SW (MSE mean = 0.01151, SD = 0.0007) are again very similar to each other in terms of how well they replicate the data. The ER network (MSE mean = 0.0124, SD = 0.00175) closely follows all the networks except FC. The ABM on FC network provides the worst fit for the validation data (MSE mean = 0.01878, SD = 0.00622).

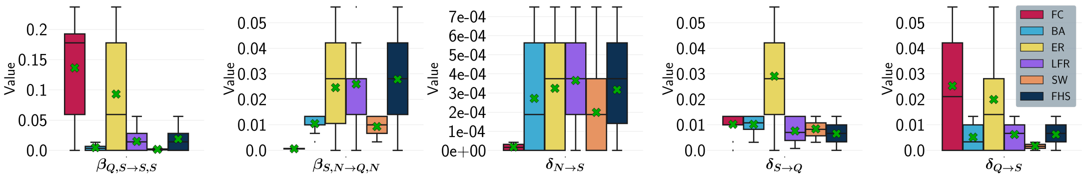
The ER and BA networks perform well in the case of the UK (almost 60% of the parameters are not significantly different from the FHS network), but in the case of the US, similarity of the parameters drops (only 40% are not significantly different). However, the LFR network performs decently in the UK data set (60% are not significantly different) and very well in the US case (none of the parameters is significantly different).

Our results thus indicate that when the average degree is kept constant, the parameter values found for the FHS network closely align to that of the LFR, ER and BA networks.

Comparing ABM calibrations between the US and UK reveals varying levels of robustness across different network types. In the LFR, SW, and FHS networks, 20% of parameters showed no significant difference (at $\alpha = 0.05$), while the ER network demonstrated greater robustness with 40% of parameters not significantly different. These four network types (LFR, SW, FHS, and ER) exhibit some resilience to geographic variations. In contrast, the BA and FC networks showed significant differences in all parameter values between the US and UK calibrations, indicating no robustness to geographic changes. This analysis suggests that ABMs calibrated



(a) UK data



(b) US data

Figure 2.5: Box plots representing the range of values for each parameter in the 100 simulations that best fit the calibration data for each network. The green crosses in each box show the mean value of the parameters. From left to right, we have the FC (red), BA (light blue), ER (yellow), FHS (dark blue), LFR (purple) and SW (orange) networks for each plot.

on FHS, ER, SW, and LFR networks maintain some consistency across different geographic contexts, whereas those on BA and FC networks are more sensitive to geographic variations.

Implications to policy When the underlying network structure is changed, the parameters best replicating the empirical data also change. Therefore, when models are used to develop policies, parameter estimation becomes very important to predict the outcome of potential new policies. If the wrong network structure is used for the model, the calibrated parameters will also be different, which would lead to inaccurate strategies being developed. However, our results (robustness of the FHS, ER, SW and LFR networks to changes in geographic regions and the similarity of parameters of LFR, ER and BA within each region) suggest that when the real-world network structure is not available, the LFR and ER networks provide a satisfactory approximation.

Thus, LFR or ER networks could potentially be used to develop strategies for controlling smoking behaviour when the local population's underlying network structure is unavailable.

2.4 Discussions

Using our ABM, we study the effect of network topology on the dynamics of the spread of smoking and see that the network structure affects it. The effect of network topology can be clearly seen when comparing results from an FC network to those obtained on other networks. However, the difference is minor within other networks when the average degree is similar.

Our UK and US results suggest that ER and LFR networks replicate the empirical data better than the other synthetic networks (that is, excluding the FHS network). Apart from this, by analysing the parameter values found during calibration, we observe that only ER and LFR networks are robust to changes in the geographic region and also return a combination of parameters of which at least 40% are not significantly different from those obtained by calibrating the model on the FHS network. Upon closer observation of the network characteristics of LFR, ER and FHS networks (Table 2.1), we see that all three of the networks have very similar average degrees ($< k > \approx 3$) and thus also form a similar number of edges. At the same time, the BA and SW have the same average degree as each other ($< k > \approx 4$) , which is a bit higher than the other networks. This suggests that the average degree might play a crucial role in the dynamics of smoking behaviour and possibly, any synthetic network with an average degree similar to that of the real-world network can be used to model smoking behaviour.

We have showed the ABM on networks can be used to reliably model social contagion trends in tobacco use dynamics. Even with the wrong network, these models can predict the trajectory of the populations qualitatively, albeit with a lower accuracy. Nevertheless, when a network changes, the parameter values returned on calibration also change. The differences in the model and its best-fit parameters become important when policymakers use the predictions to develop strategies to curb smoking. Incorrect approximation of network structure and thus the model parameters can potentially lead to the development of ineffective policies.However, our results show that most networks with similar degree distributions to the empirical FHS network yield comparable parameter values, suggesting some model robustness, though it's important to recognise that even well-calibrated models may not fully capture the complexities of rapidly changing policy environments. To enhance generalisability, policymakers should view these models as tools for generating insights and exploring scenarios, rather than as precise predictors. Combining model outputs with expert knowledge, considering multiple network structures, and regularly updating parameters based on emerging data can help create more adaptable and effective tobacco control policies in evolving societal contexts.

Our results suggest that in cases where the real-world network information is not entirely available due to practical constraints, policymakers can use LFR networks and ER networks to approximate the real-world network topology, as our social contagion process on these networks replicates the empirical data with good accuracy and, during calibration, returns parameter values which are not significantly different from the real-world network. This potentially opens the way to population-wide tobacco control interventions that exploit social contagion and will require minimal knowledge of real-world parameters, such as average number of close contacts.

ABM is a good technique for incorporating network structure and interactions between individuals to study the macroscopic outcome. Our model shows that the network structure of the population is essential while modelling smoking, which should be taken into consideration while developing policies. However, some limitations due to modelling choices to preserve the model's simplicity should be noted. A significant limitation of our study is the model's poor fit to empirical data, both within the calibration period and especially during validation. This discrepancy, consistent across all network topologies tested, suggests that our model does not completely represent the processes of smoking uptake and cessation. One contributing factor is our use of fixed coefficients over the 30-year calibration period, despite significant changes in the smoking context in both the UK and the US during this time. Factors such as evolving smoking regulations, advertising policies, and pricing have influenced smoking behaviour (Gilpin and Pierce, 1997; Millett et al., 2011; Schneider et al., 2011; World Health Organization) which our model does not completely capture. While our approach provides a baseline for comparing the effect of network structures on smoking dynamics, it simplifies the complex, time-varying nature of smoking dynamics. This simplification limits the reliability of our modelled outcomes and their applicability to real-world scenarios. Future work could address this by incorporating time-dependent parameters, segmenting the data into shorter periods with different coefficient sets, or including additional data on policy changes and other relevant factors as they become available.

Further, we have assumed that every individual behaves in the same way within a group. However, this is not the case in a real-world setting. Additionally, many social ties manifest asymmetrically, whereas we only consider undirected networks. Incorporating directionality could improve accuracy by allowing more characterisation of the uneven influence in social ties and capturing the flow of behavioural influence better. Exploring simplicial complex frameworks constitutes a promising avenue for enhancement. Such higher-order representations can capture empirical group social contagion effects (Iacopini et al., 2019).

Additionally, we have not considered vital dynamics in the model and the age-dependent nature of reaction to influence. Usually, a constant mortality rate and birth rate are incorporated to model vital dynamics. However, in network models, there is a risk of older agents gaining more centrality just because of the implementation of network growth. A careful understanding of age-dependent mortality rates and social ties should be incorporated into the model to circumvent this problem. However, this is beyond the scope of this thesis and will be explored in future work. Therefore, to limit the effects of not including vital dynamics into the model, we calibrate and validate the model in time-periods of 25-30 years.

Further, the degree of influence on smoking behaviour changes with the social tie you have with the other person (Christakis and Fowler, 2008). As a starting point in modelling influence on smoking behaviour, we have assumed that every kind of relationship affects the smoking behaviour similarly. Further, many social ties can be one-directional (as perceived by one individual in a social tie). However, as mentioned above, we have not considered any directed graphs in the model.

Since our model assumes that the total population is constant, $N = 1 - Q - S$, no new N-agents are being introduced to the system. Additionally, the S curve is calibrated against a decreasing trend observed in the empirical data, which explains the absence of variation never-smoker population.

Additionally, in a real world setting, smokers tend to be show interesting group behaviours, wherein they are part of smaller subgroups than non-smokers, and groups of smokers tend quit smoking together (Christakis and Fowler, 2008). However, due to limitations on data, we have only studied the effect of random initiation of smokers in the model. We leave this task for the future with a hope for more detailed network data on smoking.

2.5 Conclusions

We have developed an agent-based model for smoking dynamics that considers the contagious nature of smoking behaviour by including network effects. This model can act as a test-bed for network based policies and strategies to control the spread of smoking behaviour.

Our results suggest that, when interactions between individuals are used to model population-level smoking dynamics, the underlying network of the local population becomes very important. By changing the network topology from a fully-connected network to other theoretical networks and, finally, a real-world network, we show that the dynamics deviate drastically from those of a traditional ODE model.

We show that our model is robust and consistent with the historical trends observed in two countries - US and UK. In both countries, the network based on a real-world local population best replicates the trends observed.

Moreover, our results suggest that the network topology, the average number of close social ties, and the presence of communities in the population improve the accuracy of the model.

Importantly, given the difficulties in collecting data on offline social networks, we find that the LFR and ER network replicates the empirical data with accuracy and also their calibrated parameter values are not significantly different from those of the FHS network, suggesting that the LFR and ER networks can be used for social contagion models of tobacco use.

Chapter 3

Network Interventions

3.1 Introduction

Improving smoking cessation rates is one of the most cost-effective strategies for reducing preventable deaths, diseases, and disabilities on a global scale. The benefits of cessation become apparent rapidly, including a significant reduction in the risk of cardiovascular diseases, decreased risk of lung cancer and other respiratory diseases, and improved life expectancy, as highlighted in recent research (Jha and Peto, 2014; Pirie et al., 2013). Successful smoking cessation yields a multitude of advantages, significantly enhancing individual health outcomes by improving mental health (Wu et al., 2023), reducing the risk of cardiovascular diseases (Duncan et al., 2019; Mons et al., 2015), lowering the risk of cancer (Choi et al., 2018), increasing life expectancy (Taylor Jr et al., 2002), and alleviating a substantial portion of the public health burden.

Over recent decades, global smoking rates have declined considerably through collaborative efforts by international organisations, national and local governments, and non-profits (Flor et al., 2021). This progress is largely attributable to policy initiatives and frameworks like the WHO Framework Convention on Tobacco Control (World Health Organization and others, 2003). Together with the MPOWER package (Organization et al., 2008) has contributed significantly to the promotion of smoking cessation.

While these strategies have contributed to a reduction in smoking prevalence, the rate of decline is insufficient. Smoking persists as a major public health issue, with millions still using tobacco products worldwide. Furthermore, smoking prevalence is unevenly distributed across socioeconomic groups, with more deprived communities bearing a disproportionate burden (ASH Scotland, 2014). This health disparity compounds existing social inequities, as those in lower socioeconomic strata not only face higher likelihood of smoking but also greater difficulty quitting due to factors like stress, inadequate healthcare access, and tobacco marketing exposure. Individuals of lower socioeconomic status consequently suffer greater susceptibility to the adverse health consequences of tobacco use.

Many existing policies intended to reduce smoking prevalence inadvertently worsen socio-economic disparities. For example, tobacco excise taxes, while effective in motivating individuals with lower socio-economic status (SES) to attempt quitting (Hiscock et al., 2012; Pisinger et al., 2011; Vangeli and West, 2008), may inadvertently impose a disproportionate burden on those with lower incomes who are already financially strained (Blakely and Gartner, 2019; Guillaumier et al., 2015; Hoek and Smith, 2016).

Amongst the key contributors to advancement in smoking cessation are smoking cessation clinics. Individual-level cessation support offered by smoking cessation clinics has demonstrated its effectiveness in mitigating health inequalities stemming from smoking (Bauld et al., 2007; Hiscock and Bauld, 2013). Through evidence based interventions like behavioural support and nicotine replacement therapy (Bauld et al., 2010; Brose et al., 2011; Dobbie et al., 2015), these clinics aid individuals who smoke in quitting and improving their health (Stead et al., 2016).

However, the prevailing approach in most cessation clinics involves the voluntary enrolment of individuals at random who are motivated to quit for various reasons, such as health awareness, familial influence, or peer pressure (Martins et al., 2021). This approach effectively represents a form of random targeting. Despite its advantages, the random targeting approach has its limitations, as evidenced by the persistent challenges of high smoking rates and unequal declines in prevalence. Improving the targeting strategies utilised by cessation clinics is critical, especially for engaging disadvantaged populations (Jarvis and Wardle, 1999; Murray et al., 2009). Refining approaches to access and assist a more inclusive, representative smoker demographic can enhance clinic efficacy, mitigate disparities, and promote greater equity in cessation outcomes.

While the contagious nature of smoking behaviour has been extensively studied in recent years, its application in the context of tobacco control remains largely unexplored, except for the promotion of quitting campaigns via online social networks (Brown et al., 2014a; Troelstra et al., 2019). An individual's smoking behaviour, including smoking initiation, cessation, and relapse, can be significantly influenced by their immediate social network, which includes close friends and family members (Blok et al., 2017; Christakis and Fowler, 2008; Ennett et al., 2008; Go et al., 2010). That is, the chance of smoking initiation increases if your close network friend smokes, similarly the chance of an individual quitting smoking increases if their close network friend quits. By leveraging this behaviour of individuals influencing each other's smoking behaviour, we can develop effective targeting strategies for tobacco control. These strategies, in turn, have the potential to enhance the outreach of smoking cessation clinics, enabling a greater number of individuals to quit smoking while making efficient use of the existing clinic resources.

Similar network-based interventions have been extensively used in addressing various other public health issues (Hunter et al., 2019; Latkin and Knowlton, 2015) like obesity (Ashrafian et al., 2014) and to reduce HIV risk behaviours (Hoffman et al., 2013; Latkin et al., 2013), as well as towards more social phenomena like spread of opinions and beliefs (Valente, 2012) and even to promote water consumption (Smit et al., 2021). These network based interventions have exhibited remarkable success, particularly in exploiting the contagious behaviour to drive significant benefits. This has assisted in controlling the transmission of diseases, promoting healthy behaviours, and reducing the prevalence of such public health challenges. Applying similar principles to tobacco control could offer a promising avenue for curbing the global tobacco epidemic equitably.

In this chapter, we introduce a simple heuristic based on the Friendship Paradox and develop three distinct targeting strategies for smoking cessation. Using an empirically validated agent-based model, we simulate the impact of our method and demonstrate its effectiveness. Our simulation results suggest that our method outperforms random targeting, making it more effective in reducing smoking prevalence. Importantly, it significantly reduces socio-economic disparities in smoking decline, contributing to a more equitable outcome. This approach is not only effective but also cost-efficient, requiring no additional financial resources for implementation, computational resources from the cessation clinic staff, or technical training. Being a simple heuristic that enhances the current method employed, it can be employed off the shelf.

3.2 Methods

In this study, we employ the Tobacco Social Contagion Model (TSCM) as outlined in Chapter 2 to test our proposed strategies. The TSCM, being an agent-based model, has been purposefully tailored for evaluating network-based interventions.

Within the framework of this model, individuals are categorised as agents falling into one of three states: those who currently smoke tobacco products (S), those who have quit smoking (Q), or those who have never smoked tobacco products (N). The primary objective of the TSCM is to simulate long-term effects on tobacco usage dynamics by accounting for different possible state change processes. The model encompasses three primary state change processes, specifically, smoking initiation, smoking cessation, and smoking relapse. The model capture both interactions between agents and external influences to model these three processes. The model provides a unique and robust framework and has already been validated using data from the US and the UK on different types of network structure. Furthermore, TCSM can be easily adapted to different geographical regions by re-calibrating the initial prevalence of S and Q agents and, if necessary, adjusting the interaction parameters.

In this work, we have re-calibrated the model to capture smoking dynamics in Scotland from 2014-2022. As in Chapter 2, we perform a comprehensive parameter sweep to determine the parameters that simulates the closest fit to empirical data. To achieve this, we employ the Mean Squared Error (MSE) as our metric of choice. Specifically, we calculated the MSE between the modelled and empirical values for both the S and Q populations over time. The parameter set that yielded the lowest combined MSE was selected as the best-fit parameters that replicated the real-world dynamics in Scotland.

3.2.1 Model environment

Synthetic networks such as the LFR (Lancichinetti-Fortunato-Radicchi) and ER (Erdős-Rényi) can effectively simulate smoking dynamics, provided the average degree of the network remains comparable (see Chapter 2). We chose the LFR network model for our experiments because it captures the community structures found in real smoking behaviour networks.

To incorporate variations in smoking prevalence associated with socioeconomic deprivation, we use the Scottish Index of Multiple Deprivation (SIMD) (Scottish Government, 2020). The SIMD serves as a comparative measure for the whole of Scotland, dividing the country into 6,976 smaller units known as data zones, each containing approximately 500 to 1,000 residents.. This index offers a robust indicator of deprivation, considering seven distinct dimensions: (1) Income: Measures the proportion of people experiencing income deprivation. (2) Employment: Assesses working-age people experiencing employment deprivation. (3) Education: Evaluates school attendance, attainment, and working-age people with no qualifications. (4)Health: Considers factors such as mortality rates, hospital stays related to alcohol and drug misuse, and low birth weight. (5)Access to Services: Measures average drive times to key services and public transport travel times. (6) Crime: Analyses recorded crimes per 10,000 population. (7) Housing: Assesses overcrowding and lack of central heating.

By integrating these diverse factors, the SIMD provides a realistic indicator of deprivation. This approach allows for a more accurate representation of socioeconomic disparities across different regions of Scotland.

Our analysis specifically focuses on smoking prevalence across SIMD quintiles, which range from the most to the least deprived areas. This allows for a systematic examination of the impact of socio-economic disparities on smoking prevalence. It is worth noting the substantial variability in smoking prevalence among these deprivation groups. Specifically, in 2021, the most deprived quintile had a smoking prevalence rate of 24%, in stark contrast to the least deprived quintile, where the rate was only 5% (Scottish Government, 2020).

As previously discussed, the LFR network inherently possesses community structures. However, for the purpose of our experiments, we need to partition the entire population into five roughly equal-sized groups. To accomplish this, we randomly assign each of the smaller LFR communities to one of five larger groups, ensuring that the resulting groups have approximately equal sizes. Each group is initialised with a different smoking prevalence rate, reflecting the varying tobacco use patterns observed across different SIMD quintiles. Consequently, each agent in the network is assigned to an SIMD quintile, which influences their initial smoking status and susceptibility to smoking-related behaviours.

3.2.2 Network interventions

Network-based intervention strategies have been employed for addressing various social contagions in the past. These strategies often involve a combination of tactics, including individual targeting, segmentation (grouping), induction (promoting interactions), and alterations (Valente, 2012). Individual targeting focuses on identifying and intervening with the most influential individuals within the network based on network metrics like centrality measures. Segmentation involves dividing the network into distinct subgroups or communities to tailor interventions to specific needs and dynamics. In induction, connections between individuals or groups are created or strengthened to facilitate the spread of positive behaviours. Alterations involve making structural changes to the network itself, such as adding or removing connections or modifying the strength of existing ties, to reshape the network to promote desired outcomes. The effectiveness of these tactics often relies on understanding the network structure and the nature of the social contagion being addressed (Valente, 2012).

In alignment with these well-established strategies, our approach places a central emphasis on the identification of the network's most influential individuals. Leveraging the social contagion dynamics of tobacco use, we select a pre-determined number of highly influential people as our target group. Our intervention strategy involves converting these identified individuals into either "zealot quitters" or "zealot never smokers", depending on their smoking history. Specifically, those who are currently smokers or former smokers become zealot quitters, while those who have never smoked are transformed into zealot never smokers. By labelling them as "zealots," we signify that these agents will remain in their designated state throughout the simulation, regardless of any subsequent influences or interactions.

Our strategy assumes a 100% initial intervention success rate for the targeted group. However, this does not reflect real-world smoking cessation clinics, which do not achieve 100% success when targeting specific individuals. While this assumption may seem idealistic, our primary goal is to maximise the network effect of the intervention within the simulation. Some simplification is necessary to enable a comprehensive understanding of our targeted approach's potential impact. Moreover, since we apply this assumption evenly when comparing our targeted strategy to existing strategies, it allows a fair relative comparison of effectiveness between approaches. However, future work should examine how more realistic, lower success rates may influence outcomes for intervention.

Our approach aims to improve the effectiveness of the existing targeting strategy by leveraging network structure properties. The current strategy is essentially a form of random targeting, where individuals contact the stop smoking service due to medical advice, through advertisements or deciding to quit independently. However, these individuals may not have substantial influence within their local network or community. To enhance this strategy, we develop a simple heuristic based on the friendship paradox.

Friendship paradox based heuristic

The Friendship paradox is a social phenomenon first observed by sociologists in the early 1990s (Feld, 1991), and it has subsequently become a subject of interest in network science research. It states that, on average, an individual's friends tend to have more friends than the individual themselves. This property has been found not only in friendships but also more broadly across human social networks, making it a salient network characteristic (Cantwell et al., 2021).

In a pure implementation of the Friendship Paradox (FP), one would start by identifying a random smoker agent. Subsequently, a random smoker network neighbour of this initial agent would be identified. The focus would then solely be on this network neighbour, ignoring the initial random smoker. However, such an approach might not be practical for our purposes as it would mean turning away the initial smoker who has already expressed interest in quitting - an inefficient strategy. Therefore, we employ a variation of the FP paradox, where we concentrate on both the initial random smoker and their network neighbour.

Our FP-heuristic involves two phases. The initial phase uses random selection to target individuals. We then augment this by proactively encouraging these randomly selected individuals to recruit friends into the intervention. For example, rather than solely targeting 100 people randomly, we randomly select 50 individuals and have each recruit 1 additional friend. These friends can be current smokers, former smokers, or never smokers. This approach expands the intervention's reach while diversifying the targeted group to include a wider range of network members. Engaging individuals through their social ties allows us to access a broader spectrum of the population.

We have explored three different implementations of such an amalgamation of random targeting and the Friendship Paradox (FP) method. These implementations of FP heuristic are designed to be applicable to different focus groups, or sub-populations, within a larger network. These sub-populations are used to target the interventions to SIMD quintiles.

1. First Come, First Served Focus (FCFS): In the FCFS implementation, the first step is to identify a random smoker agent from within the focus group. This se-

lected agent is immediately added to the target group for intervention. Following this, the algorithm examines the agent's network neighbours to identify smokers. If such smoker neighbours exist, one of them is chosen at random and also added to the target group. This strategy focuses on quickly identifying and targeting the first smokers, along with one of their smoker neighbours if available. The approach used is described in detail in Algorithm 1. This approach unlike a pure implementation of the FP paradox addresses the practical concern of not turning away individuals who have already been identified and may be ready to quit.

2. Pairs First: The Pairs First implementation begins by selecting a random smoker agent from the focus group. The algorithm then checks if this initial agent has any smoker neighbours. If such neighbours exist, one is chosen at random and both the initial agent and the randomly selected neighbour are added to the target group. If no such neighbour exists, the initial agent is ignored, and another random smoker is selected. The process continues until there are no more smokers in the focus group with at least one smoker neighbour. Subsequently, additional random smoker agents are added to the target group to fill any remaining slots. This approach maximises the effect of social influence. The algorithmic implementation of the FP Pairs First strategy is described in detail in Algorithm 2

3. Pairs First Restricted: The Pairs First Restricted implementation refines the Pairs First approach by adding an additional criterion for selection. After identifying a random smoker agent from a designated group, the algorithm checks whether this agent has a smoker neighbour who is also a part of the same focus group. If such a neighbour exists, both the agent and the neighbour are included in the target group for intervention. If not, the initially selected agent is ignored and a new agent is chosen for examination. The process continues until no more pairs that fit the criteria are found in the focus group. Following this, random smoker agents are added to the target group to meet the required number of participants. We try to address the potential disparities in smoking cessation

Algorithm 1 FP FCFS Implementation

```
Input: network, no_of_target, group_members
Initialise an empty target list
smoker_in_group $\leftarrow$ get_smokers(group_members)
while len(target) $\leq$ no_of_target do
    random_smoker $\leftarrow$ random.choice(smoker_in_group)
    Remove random_smoker from smoker_in_group
    Add random_smoker to target if not already in target
    neighbour_nodes $\leftarrow$ network.get_neighbours(random_smoker)
    smoker_neighbour_nodes $\leftarrow$ filter_smokers(neighbour_nodes)
    if smoker_neighbour_nodes then
        friend $\leftarrow$ random.choice(smoker_neighbour_nodes)
        Add friend target if not already in target
    end if
end while
Activate intervention on agents in target
```

across different subgroups through this approach. The algorithm used for the FP Pairs First (Restricted) strategy is detailed in Algorithm 3.

3.2.3 Experimental strategies

We implement our FP-heuristic across three distinct intervention focus groups (sub-populations), developed in consideration of substantial socioeconomic disparities in smoking prevalence across SIMD quintiles

1. **ST1** : Targeting the Entire Population: We apply our targeting heuristic across the entire population, irrespective of SIMD quintile, to assess its broad impact.

2. **ST2** : Targeting the Most Deprived SIMD quintile: We focus solely on the most deprived SIMD quintile to assess the impact of our targeting heuristic on reducing socioeconomic disparities in smoking prevalence.

3. **ST3** : Focusing on the Two Most Deprived SIMD quintiles: We focus on the two most deprived quintiles, reaching a substantial number of disadvantaged population.

Algorithm 2 FP Pairs First Implementation

```
Input: network, no_of_target, group_members
Initialise an empty target list
smoker_in_group $\leftarrow$ get_smokers(group_members)
while len(target) $\leq$ no_of_target do
    random_smoker $\leftarrow$ random.choice(smoker_in_group)
    Remove random_smoker from smoker_in_group
    neighbour_nodes $\leftarrow$ network.get_neighbours(random_smoker)
    smoker_neighbour_nodes $\leftarrow$ filter_smokers(neighbour_nodes)
    if smoker_neighbour_nodes then
        friend $\leftarrow$ random.choice(smoker_neighbour_nodes)
        Add friend and random_smoker to target if not already in target
    end if
    if len(smoker_in_group) == 0 then
        smoker_list $\leftarrow$ list of smokers in group_members not in target
        random_smoker $\leftarrow$ random.choice(smoker_list)
        Add random_smoker to target if not already in target
    end if
end while
Activate intervention on agents in target
```

These distinct applications allow comparative analysis of the targeting heuristic's impact on population segments with varying deprivation levels and associated smoking patterns.

In the supplementary section A.6, we demonstrate that our model is robust to variations in population size. However, for the experiments conducted in this chapter, we have increased the agent population from 1,000, as indicated in Chapter 2, to 10,000 (while keeping computational requirements low). This adjustment was made to ensure that each SIMD group contains a sufficient number of agents. Additionally, the increase in agent population ensures that the target demographic does not become overly limited when we test our interventions.

The interventions were simulated using the agent-based modelling framework MESA in python (Kazil et al., 2020).

Algorithm 3 FP Pairs First (Restricted) Implementation

```
Input: network, no_of_target, group_members
Initialise an empty target list
smoker_in_group ← get_smokers(group_members)
while len(target) $\leq$ no_of_target do
    random_smoker ← random.choice(smoker_in_group)
    Remove random_smoker from smoker_in_group
    neighbour_nodes ← network.get_neighbours(random_smoker)
    smoker_neighbour_nodes ← filter_smokers(neighbour_nodes)
    smoker_neigh_in_group ← filter_nodes_ingroup(smoker_neighbour_nodes)
    if smoker_neigh_in_group then
        friend ← random.choice(smoker_neigh_in_group)
        Add friend and random_smoker to target if not already in target
    end if
    if len(smoker_in_group) == 0 then
        smoker_list ← list of smokers in group_members not in target
        random_smoker ← random.choice(smoker_list)
        Add random_smoker to target if not already in target
    end if
end while
Activate intervention on agents in target
```

3.2.4 Evaluation metrics

To assess the impact of our interventions, we employ two primary metrics: one focused on prevalence and the other on measuring inequality.

Relative change in prevalence: We compute the relative change in prevalence 10 years post-intervention for each implementation and experimental strategy. This is compared with the prevalence from a random targeting strategy applied at the whole population level (ST1).

Inequality in prevalence: To evaluate each targeting method and strategy's impact on socioeconomic disparities, we examined the absolute inequality in smoking prevalence between the most and least deprived SIMD quintiles. Specifically, we calculated the difference in prevalence between the two most deprived quintile and two least deprived (Prevalence in the two most deprived - Prevalence in the two least deprived). This quantified how each approach affected existing disparities across socioeconomic groups and determined if the current inequality was reduced.

3.2.5 Data

Our study incorporates official smoking prevalence data obtained from Scottish public health data sources (Scottish Government, 2020). We use the prevalence of smokers and quitters in Scotland to calibrate our model.

3.3 Results

We start our analysis by comparing the effectiveness of the Friendship Paradox (FP) based heuristic in reducing smoking prevalence relative to random targeting employed at the whole population level. We explore three different implementations of the FP strategy, which include targeting the entire population (ST1), focusing solely on the most deprived group (ST2), or extending to the two most deprived groups (ST3). In each case, we compare the outcomes to those of random targeting across the entire population.

Next, we examine how these interventions impact socio-economic disparities in smoking prevalence. Our objective is to determine the combination which is more effective in both reducing overall prevalence and diminishing these inequalities.

3.3.1 Impact on prevalence

We assess the prevalence of smoking 10-years post intervention to compare the effectiveness of the FP-based heuristic using all three implementation strategies.

We did a marginal gains analysis for each strategy to analyse our results. The results clearly indicated a significant improvement in each of the FP-based heuristic strategies when compared to the existing random targeting strategy, as shown in Table 3.1. Specifically,

In ST1, the FP FCFS implementation outperforms the Pairs-First approach in reducing smoking prevalence. The FP FCFS and Pairs-First strategies, when applied to the entire population, yielded statistically significant marginal gains of 3.74% and 1.97%, respectively, compared to random targeting.

When targeting the most deprived SIMD group (ST2), FP Pairs-First implementation resulted in a substantial marginal gain of 2.45%, while the FP FCFS approach only had a gain of 0.42%, this gain was not statistically significant at $\alpha = 0.05$. In contrast, the random targeting method in ST2 did not show any statistically significant gain (0.05%).

Finally, when we focus on the most deprived and second most deprived groups, the FP FCFS and FP Pairs-First strategies produced statistically significant gain of 2.26% and 2.14% (statistically significant even at $\alpha = 0.01$), respectively.

The FP Pairs First (restricted) approach for both ST2 and ST3 strategies did not have a significant improvement over random targeting (ST1). This lack of advantage is primarily due to the constraints imposed by the 'restricted' conditions, which resulted in a low number of targeted pairs of smokers within the focus group. On average, the FP Pairs First strategy targeted 17 to 21 times as many pairs of smokers as the restricted version, while FP FCFS targeted between 7 to 9 times as many. Furthermore, these restrictions dampen the network effect. A similar effect can also be observed in the case of random targeting when applied to ST2 and ST3. In contrast, FP methods without such restrictions tend to amplify the network effect, especially as the pairs can involve individuals from different SIMD groups.

Heuristic	Whole Pop (ST1)	SIMD1 (ST2)	SIMD1 & SIMD2 (ST3)
Random	-	0.05%	0.72%
FP FCFS	3.74%	0.42%	2.26%
FP Pairs First	1.97%	2.45%	2.14%
FP Pairs First (Restricted)	–	1.21%	0.33%

Table 3.1: Marginal gains in prevalence reduction after 10 years for each heuristic and strategy combination, compared to random targeting of the whole population (ST1). Results are based on 1000 iterations with randomly grouped communities.

3.3.2 Impact on absolute inequality

Table 3.2 summarises the absolute difference after 10 years of implementing the strategy. Among all the cases tested, the heuristics applied using ST1 had the maximum absolute inequality, with the random strategy (the current existing one) performing the worst in this context.

Among the FP approaches without restrictions, the FP Pairs First (ST2) achieved the best results with a relative reduction of 11.71% in absolute inequality compared to random targeting (ST1), while FP Pairs First (ST3) closely followed with a 10.67% relative reduction. The FP Pairs First method also outperformed the current random targeting strategy by 9.58% and 8.71% when applied to ST2 and ST3, respectively.

As expected, the random targeting method, when applied to ST2 (resulting in an 11.79% relative drop) and ST3 (with a 12.17% relative drop), along with FP Pairs First (restricted) in ST2 (yielding a 12.68% relative drop) and ST3 (achieving an 11.77% relative drop), demonstrated its effectiveness in reducing inequality. The same factors that limit the impact of prevalence reduction for these methods also contribute to a more pronounced reduction in inequality. Since the selected targets are concentrated within a smaller population, and there are insufficient pairs of individuals chosen, the dampening of the network effect, as discussed earlier, ensures that the prevalence only decreases within the specific SIMD group that is the focus In contrast, the FP methods without restrictions not only decrease the prevalence within the SIMD group almost equivalently but also extend the reduction to the entire population due to pairs of smokers being selected across the population.

Heuristic	Whole Pop (ST1)	SIMD1 (ST2)	SIMD1 & SIMD2 (ST3)
Random	6.93	6.11	6.08
FP FCFS	6.64	6.19	6.12
FP Pairs First	6.75	6.26	6.32
FP Pairs First (Restricted)	–	6.05	6.11

Table 3.2: Table shows the absolute inequality in smoking prevalence between the two most deprived SIMD vs. two least deprived SIMD quintiles, 10 years after the intervention when communities are grouped randomly. We can see that the FP Pairs-First (restricted) method for ST3 intervention has the lowest absolute inequality, while all of the heuristics in ST1 are comparatively worse than others. Among all heuristics applied to ST1, the Random targeting (the current standard) reduces the inequality the least.

3.4 Discussions

From the analysis presented in Section 3.3 and as illustrated in Table 3.1, it becomes evident that our approach, leveraging the FP-based heuristic (FP FCFS and FP Pairs first), consistently outperforms the current existing strategy of random targeting. While the reported marginal gains for each implementation may appear as modest improvements individually, the true significance emerges when we scale these numbers to the entire population. The cumulative impact of these improvements on a population-wide scale is substantial and underscores the value of our approach in achieving reductions in smoking prevalence.

While a pure implementation of the Friendship Paradox (FP) would involve exclusively targeting the friends of the initial random agent who smokes, this pure FP method may not be practical in real-world applications. This impracticality arises from the exclusion of individuals who are motivated to quit smoking and focusing solely on the recruitment of their friends.

The FP heuristic implementation discussed in Section 3.3 represents a more efficient and balanced approach in terms of practicality. In this implementation, we target both the initial random smoker and their smoking friends in pairs thus using an amalgamation of the random and the pure FP heuristic. This setup capitalises on the fact that influences within an agent's local social network are also addressed. Consequently, this approach enhances the network effect while mitigating the impracticality associated with exclusively focusing on the friends of the initial smoker.

The two different implementations were tested, since there might be instances when there are not enough pairs of individuals who smoke. Amongst the two implementations, the FP FCFS projects a more realistic and practical solution since we do not reject the initial random smoker agent even if they do not have a smoker network neighbour. This can be seen as a smoking cessation clinic supporting the motivated individuals who intend to quit and asking them to bring a smoker friend along, as compared to FP pairs first, where the smoking cessation clinic would reject the motivated individual if they do not have a smoker friend initially, and only support them, after they have exhausted all individuals who have a smoker friend.

The analysis of absolute inequality in smoking prevalence between SIMD quintiles provides further insights into the effectiveness of our targeting strategies. As shown in Table 3.2, the FP-based methods, particularly when applied to specific SIMD groups (ST2 and ST3), demonstrated a reduction in absolute inequality compared to random targeting of the whole population. Specifically, the FP Pairs First method achieved relative reductions of 11.71% and 10.67% when applied to ST2 and ST3 respectively. These results indicate that our network-based approach not only improves overall prevalence reduction but also contributes to narrowing the socioeconomic gap in smoking rates. It's worth noting that while restricted versions of our methods showed even greater reductions in inequality, they had less impact on overall prevalence, highlighting the complex trade-offs in addressing both overall smoking rates and socioeconomic disparities simultaneously.

Cost effectiveness of the FP method Our proposed targeting improvement not only reduces prevalence more equitably, but is also essentially a zero-cost method. The current system need not undergo major changes - rather, it requires only slight adjustments. For instance, instead of randomly targeting 1% of smokers, clinics could target the same proportion but with the difference that its a combination of random smokers and their smoker friends who is recruited. Such adjustments harnessing social ties can be readily implemented since they demand no additional resources. While cessation clinics are already highly cost-effective from an economic standpoint (Levy et al., 2022), our approach further leverages social contagion to enhance public health returns on investment. By tapping into social contagion, we can magnify societal benefits beyond the direct impact of intervention participants.

In similar social contagion systems, network centrality measures have served as essential tools for identifying influential individuals. Intervention methods that rely on these centrality-based approaches have proven highly effective in other public health contexts (Valente, 2012). However, employing a network centrality measure may not be practical when considering factors like computational costs, the unavailability of network data, and the technical expertise of cessation clinic staff. In contrast, the FP-based approach we propose is a simple heuristic that does not entail additional computational expenses or technical expertise. It can be readily implemented without requiring any specialised adjustments.

The FP heuristic is an interesting approach in the context of smoking cessation for effective targeting. Nevertheless, it is important to acknowledge some limitations resulting from specific modelling choices made to maintain the model's simplicity. First, our model initialises the population with smoking states at random. There is evidence to suggest that an individual's likelihood of being a smoker is significantly higher if they have close friends or family members who smoke (Saari et al., 2014). This is because the number of smoker-smoker (S-S) connections formed is much greater than smoker-non-smoker (S-N) connections. However, data at a network level or the necessary information to construct such a network is not currently available. Consequently, we employ a random initialisation setup.

Second, in our experiments, we assume that the strategy is deployed only once, targeting 1% of the population in the first year without any follow-up. In reality, such interventions are often carried out regularly. However, our model assumes that once a smoking cessation clinic targets an individual, there is a 100% success rate in them quitting, leading to a shortage of smokers to target after a few years of intervention. In reality, the success rate of smoking cessation clinics is not perfect and is much lower. Nevertheless, our model is designed to compare the effects of different strategies relative to the current standard, and we have made this assumption accordingly. We also hypothesise that the effects of our interventions will compound with each follow-up intervention.

It's worth noting that both of these limitations primarily affect the FP methods. Therefore, it is reasonable to assume that the results presented in this chapter provide a conservative estimate of the potential improvements that social contagion-based network interventions can bring to tobacco control.

3.5 Conclusion

In this work, we have demonstrated a novel approach to improve the effectiveness of smoking cessation clinics by incorporating network-based targeting strategies. Leveraging computational agent-based modelling, we evaluated a heuristic based on the friendship paradox for identifying influential individuals in the social network. Through simulation experiments, our approach outperformed the existing standard practice of random targeting across several metrics.

Specifically, we designed and tested three implementations of the friendship paradox heuristic focused on the general population as well as socio-economically disadvantaged groups. In all cases, the heuristic yielded significant marginal gains in reducing smoking prevalence over 10 years compared to random targeting. Targeting the most deprived quintile using the “pairs first” heuristic variant achieved a substantial 2.45% reduction versus standard practice.

Critically, our approach also effectively reduced socioeconomic disparities in smoking decline. Focusing on the two most deprived quintiles, the heuristic decreased absolute inequality in smoking rates between the two highest and two lowest deprivation groups. The reductions ranged from 2.5% to 11.7% lower absolute inequality compared to random targeting based on focus group strategy. This reduction in inequality is a direct result of concentrating resources where they are most needed, effectively levelling the playing field across socioeconomic groups.

Moreover, our friendship paradox strategy requires no additional resources, as it involves only minor adjustments to current clinic recruitment practices. By leveraging social contagion dynamics, our approach enhances outcomes while remaining highly cost-effective.

Chapter 4

Modelling Socioeconomic and Spatial Influences on Tobacco Use

4.1 Introduction

In previous chapters, we explored smoking cessation and discussed policies aimed at harnessing the effects of social contagion cessation to maximise the impact of cessation clinics. In this chapter we develop a spatial model which captures multiple factors involved in smoking initiation.

While encouraging smoking cessation programs is crucial, equal attention should be directed towards preventing smoking initiation and understanding the continued use of tobacco products (Nagelhout et al., 2012). Focusing on smoking cessation facilitates a rapid decrease in prevalence, but addressing smoking initiation yields a longer-term impact, ultimately making smoking obsolete (OBE, 2022). Furthermore, understanding the factors contributing to the continuation of smoking can help create the right environment to encourage quitting and abstaining from smoking.

Initiation rates among teenagers have witnessed substantial declines in the UK (Opazo Breton et al., 2022), attributed to the implementation of policies like increase in legal age for tobacco sales from 16 to 18 in Scotland (Anyanwu et al., 2020); nevertheless, there remains room for improvement especially if the targets set by different countries and international agencies for smoke-free generation (Scottish Government, 2018) need to be met.

To gain a comprehensive understanding of the dynamics around smoking initiation, it is important to navigate the complex tobacco landscape. Smoking initiation in adolescents is influenced by a multitude of factors, each contributing to the complexity of it. Initiation is generally a gradual process involving curiosity, weakened non-smoking intentions, and experimentation - all predictive of progression to regular smoking (Pierce et al., 2005). Further, evidence indicates individuals who pick up smoking later on often believe initial experimentation is low-risk (Choi et al., 2003). This highlights opportunities for early interventions, unlike cessation where nicotine dependence already affects individual's behaviour and decisions. The slow-paced nature of initiation provides multiple leverage points to deter progression from experimentation to regular smoking through targeted policies. However, in order to effectively harness these requires elucidating the multifaceted individual, social, and environmental factors that drive individuals from never-smoking to regular tobacco use.

One such influential social factor is exposure to parental smoking, a well-documented contributor to the likelihood of an offspring picking up smoking (Gilman et al., 2009). The risk of picking up smoking escalates significantly in proportion to the number of parents who smoke. Compared to children of non-smoking parents, children with at least one smoking parent are 1.72 times more likely to take up smoking. This risk further increases if both parents smoke, with the odds rising to 2.73 times higher (Leonardi-Bee et al., 2011). It is worth noting that the smoking habits of an individual's close friends and family members consistently shape their own smoking patterns and behaviours including initiation, continuation and cessation of smoking (Christakis and Fowler, 2008). Further, studies have indicated that most individuals initiate smoking before the age of 21 (Ali et al., 2020; Chassin et al., 1990; Lydon et al., 2014). This also happens to be the age when many individuals are extremely susceptible to peer influences (Steinberg and Monahan, 2007). It is, therefore, not surprising that being exposed to smokers within one's social circle is a strong predictor of smoking initiation (Harakeh and Vollebergh, 2012; Liu et al., 2017).

Similar to social factors, environmental factors, such as tobacco retailer density, have been strongly linked to smoking initiation and usage (Cantrell et al., 2015; Marsh et al., 2021; Shortt et al., 2016). While this connection is often attributed to the widespread availability of tobacco products in areas with a high retailer density, which may contribute to increased youth smoking initiation (Pokorny et al., 2003), it's important to note that the causal relationship could be bidirectional, with retailers potentially responding to existing market demand as well. Furthermore, a higher retailer density may also lead to reduced costs associated with travelling to purchase tobacco, and some retailers may also attempt to sell cheaper tobacco products (Henriksen et al., 2017; Shortt et al., 2021). It is important to note that tobacco retailers are more prevalent in socio-economically deprived areas, where both smoking prevalence and initiation rates tend to be higher (Hiscock et al., 2012). Additionally, research has shown that adolescents from socio-economically deprived regions are more exposed to tobacco retailers (Caryl et al., 2020). Therefore, understanding the effects of retailer density and the availability of tobacco is not only crucial for comprehending initiation and prevalence but essential for the development of more equitable policies.

Individual-level attributes also significantly shape the risk of smoking initiation, including factors such as age (O'Loughlin et al., 2014), lifetime cigarette consumption (indicating addiction tendencies) (Sargent et al., 2017), and financial capacity to purchase tobacco products (Cui et al., 2019). While initiating smoking at a younger age is associated with a higher risk of continued use of tobacco (Breslau and Peterson, 1996; Breslau et al., 1993), experimenting with smoking increases the likelihood of becoming a regular smoker (Breslau and Peterson, 1996; Breslau et al., 1993). An increase in pocket money has also been linked to a higher intensity of smoking and probability of initiation (Cui et al., 2019).

Individual, social, economic, and spatial factors interact in complex ways to shape the tobacco policy landscape and specifically, smoking initiation. Navigating this multifaceted system requires a complex systems approach to examine the intricate dynamics between diverse elements and assess implications holistically (Rutter et al., 2017).

Computational models of complex systems have proven to be invaluable tools for addressing various public health challenges. The COVID-19 pandemic underscored the significant impact of rigorously validated models in guiding decision-making (Aleta et al., 2020; Ferguson et al., 2020; Kerr et al., 2021; Kucharski et al., 2020). Beyond infectious disease models, similar modelling-based tools have been used to develop and evaluate policies across diverse domains. These applications encompass addressing public health challenges related to obesity (Orr et al., 2016), reducing the population burden of mental disorders (Cerdá et al., 2015) and addressing alcohol usage (Gorman et al., 2006), as well as simulating policies aimed at reducing tobacco retailer density (Luke et al., 2017).

Given the effectiveness of computational modelling in addressing other public health concerns, a comparable framework for studying and testing individual-level, spatial, and social tobacco control policies will be useful. Such a framework can improve our understanding of the complex interplay of factors influencing tobacco use and facilitate the design of evidence-based tobacco control policies.

Using an agent-based modelling framework we can develop a data-driven, empirically validated model that provides a controlled platform to gain insights into adolescent smoking initiation. Such a framework can be used to:

1. Provide an adaptable framework to study a wide range of individual-based, location-based, and population-level interventions.
2. Inform evidence-based policies by testing interventions through simulation.
3. Evaluate the equity implications of potential tobacco control policies.

In this chapter, we have developed such a comprehensive heterogeneous agent-based modelling framework to focus on smoking initiation. To establish this model, we incorporate various determinants of smoking initiation, encompassing parental influence, peer influence, addiction, retailer availability, and financial resources. We will evaluate different policies and their effects on smoking initiation, prevalence, and equity in the following chapter.

Additionally, we provide an in-depth overview of the datasets employed to inform the model in section 4.7.4. We delve into the details of how we represent diverse socio-economic scenarios within the model. Finally, we outline the process of model calibration and validation, ensuring its reliability and accuracy. Our model is adaptable and can be readily employed to assess a wide range of policies, spanning interventions targeted at individuals, location-based strategies, and tobacco-specific measures, all while taking into account socioeconomic considerations.

4.2 The tobacco town model

The Tobacco Town (TT) model, developed by Luke et al. (2017), serves as a reference point for the development of our model. This ABM was designed to examine the effects of tobacco control policies on retailer density and consumer behaviour. The TT model simulates four types of towns based on Urban / suburban classification and income levels (rich and poor). Within these environments, they model all adult smokers as agents, each with attributes including smoking rates, mode of transport,wage, home/work location. The model primarily focuses on four types of policies: random retailer density reduction, restriction of a particular type of retailers, limiting proximity of retailers to schools, and limiting proximity of retailers to each other. The TT paper studied outcome measures centred on the accessibility of tobacco products and the overall cost (including travel and time costs) for consumers with simulation running over a short-term period of 30 days.

The TT model has several strengths that make it an appropriate starting point for our research. It is an innovative application of ABM to tobacco control policy analysis incorporating spatial elements and thus considering urban and suburban environments. The inclusion of this socio-economic factors through different income layers add another layer of realism to the simulations. However, the model also has limitations. It solely focuses on adult smokers, excluding adolescents and youth and initiation dynamics. The short simulation time frames also do not capture long term effects. Moreover the model does not account for social influence or peer effects on smoking behaviour, which are crucial factors in tobacco use behaviour, especially in younger populations.

Despite these limitations, the TT model provides a good foundation for our work due to its use of ABM to simulate complex policy effects, incorporation of spatial and socio-economic factors, and focus on retailer density. To address, our specific research questions, we found it necessary to make significant modifications and extensions to the model.

In our model, we have expanded the demographic scope to include three age groups (15-18, 18-21, 21-24), capturing adolescent and youth smoking initiation that were missing in the TT model. We have also incorporated social influence, peer effects, parental influence on smoking behaviour and illegal sale of tobacco to adolescents thus incorporating various social behaviour in the model.

A crucial aspect of our modification is the adaptation to the Scottish context. While the TT model was based on US data and town types, our model uses the Scottish specific data on retailer density, socioeconomic indicators (SIMD), and urban-rural classifications to improve the models relevance to Scottish policymakers and researchers.

We have also extended the simulation time frame from the original 30 days to 10 years, allowing us to capture long-term policy effects that may not be apparent in shorter simulations. Our policy scope has also been broadened to include price increases and minimum age policies, in addition to retailer reduction strategies. Further, our outcome measures have been also been expanded beyond cost, to include tracking of consumption patterns, retailer sales and the socio-economic disparities. Additionally, we have Incorporated stakeholder engagement in to our model development process, particularly through a feedback from Public Health Scotland, ensuring out model addresses relevant policy questions.

4.3 Model description

Building upon the foundation laid by the TT model, we have developed a significantly enhanced agent-based model capturing smoking initiation behaviour. This section describes our model's structure and underlying assumptions.

4.3.1 Individuals

Our research focuses primarily on the initiation of smoking behaviour, therefore we target age groups that are most susceptible to uptake smoking. Extensive empirical evidence consistently indicates that a significant majority, approximately 70%, of individuals who begin smoking do so before reaching the age of 18 (Filippidis et al., 2015), and an even higher proportion, around 85%, initiate smoking by age 21 (Ali et al., 2020). As a result, our model is specifically oriented towards individuals in the age range of 15 to 24 years. Within this demographic, we have three distinct categories of agent types: *adult agents* (ages 21-24), *young-adult agents* (ages 18-21), and *adolescent agents* (ages 15-18). The agents within these discrete age groups employ distinct behavioural patterns.

Each agent in all three age groups is associated with a smoking status. This can be a never-smoker or a smoker. Further, in the case of *adolescent agents* and *young-adult agents*, smoker status is further divided into experimenter and regular smoker (an individual who has smoked more than 100 cigarettes in their lifetime). We model the agents in such a way that, the agents below the age of 21 (*adolescent* and *young-adult agents*) can experiment with smoking due to their peer influence, which can eventually lead to regular smoking.

Below, an overview of agents in each age group is given.

Adolescent agents All *adolescent agents* within our simulation engage in regular movement between their homes and schools. We assume that they use public transport for this. The initiation of smoking behaviour in these agents is affected by factors such as peer influence (Liu et al., 2017), parental smoking habits(Gilman et al., 2009) and addiction (Sargent et al., 2017). We make the assumption that every adolescent who attends the same school or who lives near their households possesses the potential to impact the smoking behaviour of their peers. Furthermore, the weekly pocket money allocated to each agent plays a pivotal role in determining their tobacco purchasing behaviour. This financial resource is derived as a fraction of the average salary earned by working adults within the specific neighbourhood type, as indicated by empirical data.

It is essential to note that there is a strict age restriction on the sale of tobacco products in the UK. The sale of tobacco products or procuring tobacco products to a person under 18 are both prohibited by law. Nevertheless, empirical observations highlight that, despite the existence of such legal regulations, a substantial number of adolescents manage to obtain tobacco products. They do so either through retailers who do not rigorously adhere to these laws or by enlisting the assistance of older peers for purchase (proxy-buying). Our model takes these illicit means into account for modelling the purchasing behaviour of *adolescent agents*. These behaviours are discussed in detail in Section 4.5.1

Never-smokers among the *adolescent agents* abstain from any interaction with tobacco retailers.

When they turn 18 years old, these agents become *young-adult agents*.

Young-adult agents The behaviour of *young-adult agents* in our simulation revolves around their movement between their homes and workplaces, utilising various modes of transportation such as public transit, driving, or others (can include walking, skateboards, wheelchairs etc.). Their initiation behaviour mirrors that of *adolescent agents*, where they experiment with tobacco products due to factors such as peer influence, parental smoking habits, and addiction. It is important to note that all *young-adult agents* in the model are considered peers to fellow *young-adult agents* within their local vicinity, whether in residential or work areas, and they can influence one another's smoking behaviour.

Unlike adolescents, *young-adult agents* have a fixed income since they are employed. We calculate their earnings based on their hourly wage, assuming a standard 35-hour work week, using data from the wage distribution described in 4.7.4.

When it is legally permissible for agents in the age group of 18-21 to purchase tobacco products, they follow a purchasing behaviour that involves optimising the number of cigarette packs to purchase and selecting the most suitable retailer from which to make the purchase. The specific details of this purchasing behaviour are elaborated upon in Section 4.5.1.

Similar to *adolescent agents*, never-smokers among the *young-adult agents* abstain from any interaction with tobacco retailers.

When their age becomes 21, they become *adult agents*.

Adults Much like the *young-adult agents* in our simulation, the *adult agents* also engage in routine commutes between their homes and workplaces, employing various modes of transportation such as public transit, driving, or others. However, a notable distinction in our model is that *adult agents* no longer engage in smoking experimentation. They fall into one of two categories: they are either categorised as never smokers or as regular smokers.

Similar to *young-adults*, *adult agents* receive a fixed income in the form of a salary, and we primarily use hourly wages (35 hour weeks).

For those *adult agents* classified as regular smokers within this age group, their procurement of cigarettes is guided by a utility function and a determination of the appropriate quantity of cigarette packs to purchase from a retailer. This purchasing behaviour is explained in Section 4.5.1, analogous to the approach applied to *young-adult agents*.

Conversely, never-smokers among the *adult agents* abstain from any interaction with tobacco retailers.

4.4 Environment

Our model is built upon a square lattice graph, with the environment entities distributed on the nodes of this lattice. The edges connecting these nodes represent roads, and agents traverse these roads for movement. The size of the lattice grid varies depending on the specific area being modelled.

Attribute	**Description**
unique id	Unique identifier for each agent.
home	Identifier for the agent's home location.
school/workplace id	Identifier for the agent's school location (adolescents) or workplace.
home loc	Location of the agent's home.
school/work loc	Location of the agent's school (adolescents) or workplace.
current loc	Current location of the agent, initially set to the home location.
parental influence	Fraction of parents who smoke.
money / wage	Weekly allowance or income of the agent (for adolescents class) or hourly wage.
age	Age of the agent in days.
status	Smoking status of the agent.
discount rate	Discount rate for purchasing cigarettes (for Youth and Adult class only).
smoking rate	Number of cigarettes smoked per day.
inventory	Number of cigarettes in the inventory

Table 4.1: Variables associated with individual agent types in the model, including unique identifier, locations, smoking status, age, financial attributes, and inventory.

In our experimental setup, we establish different agent neighbourhoods, considering socio-economic factors. These neighbourhoods are determined based on a combination of SIMD (Scottish Index of Multiple Deprivation) quintiles and a two-fold urban/rural classification, leading to ten distinct combinations. However, the experimental framework discussed in this chapter, along with the presented results, primarily focuses on four extreme types of agent neighbourhoods similar to the Tobacco town paper (Luke et al., 2017) to facilitate a deeper analysis. Specifically, we model the following areas:

1. Most-deprived Rural (1st SIMD quintile, Rural classification)
2. Most-deprived Urban (1st SIMD quintile, Urban classification)
3. Least-deprived Rural (5th SIMD quintile, Rural classification)
4. Least-deprived Urban (5th SIMD quintile, Urban classification)

In Scotland, most granular data is usually available at a datazone level. These are small geographical areas on which small area statistics are available. To minimise computational costs, we limit our modelling to a total of 30 datazones, allowing us to efficiently simulate and analyse these specific neighbourhood types while keeping the computational costs low.

4.4.1 Environmental entities

Within our model, we introduce four specific entities in the environment. These entities are Retailers, which serve as the outlets for sale of tobacco products, as well as locations such as Schools, Workplaces, and Homes. These locations aim to replicate the daily routines of individuals from different age groups.

Household, schools and workplaces Each individual agent in the model is linked to an environmental agent representing their household, as well as a Workplace/School (Secondary schools) agent, which designates the primary locations where individual agents spend the majority of their time. The initial placement of the respective entities are determined through a random allocation process based on density parameters derived from empirical data.

In our model, the retailers play an active role within the model, actively engaging with the agents for the sale of tobacco products. Further details on the role and behaviour of Retailers are elaborated upon below.

Retailers The Retailer agent in our model represents a diverse range of tobacco retail outlets. It is well-established that various types of retailers sell tobacco products, and their characteristics, including outlet density, product prices, and compliance with age verification for sales to minors, can differ significantly. Consequently, we model distinct types of retailers, each with its own unique interactions and behaviours influenced by these factors.

In our model, we account for different types of retailers based on the categorisation by the Chartered Trading Standards Institute (Chartered Trading Standards Institute 2015, 2015). These retailers are primarily grouped into eight distinct categories, with varying probabilities of selling cigarettes to minors (see section 4.7.4 for more details):

1. Off-licence (premises selling alcohol for off-site consumption)
2. Large retailers (national supermarket chains)
3. Newsagent (shops selling newspapers, magazines, and other items)
4. Petrol station (shops attached to fuel stations)
5. Pub/club (on-licensed premises)
6. Private home (domestic dwellings)
7. Small retailers (independent or smaller chain stores)
8. Others (various outlets not fitting other categories)

These variations among retailer types allow us to create a more comprehensive and realistic representation of the tobacco retail landscape within our model. It enables us to study how different retail environments can influence tobacco purchasing behaviours and the accessibility of tobacco products to individuals of different ages.

4.5 Interactions and behaviours

4.5.1 Purchasing behaviour

The purchase behaviour within our model is divided into two distinct phases: the decision to buy cigarettes and the actual process involved in making the purchase.

Purchase decision Every time-step, in the morning, each agent (who are regular smokers) within our model makes a decision regarding whether to buy or attempt to buy tobacco products. This decision is based on a straightforward criterion: if the number of cigarettes in their inventory falls below their daily smoking rate, they choose to purchase cigarettes.

As previously mentioned, all agents within our model, irrespective of their age group, engage in the purchase of tobacco products, either directly from retailers or indirectly through other agents. Our model tries to capture these behaviours by implementing distinct purchasing behaviours, contingent upon whether the agent's age falls within the legal requirements for procuring tobacco products and kind of income (pocket money or salary).

When agent's age is in the legal bracket - adult and young-adult agents

In cases where an agent's age falls within the legal range for purchasing tobacco products, we adopt the purchase behaviour utility function from the approach utilised in the "Tobacco Town" paper (Luke et al., 2017), with some modifications to suit our model's context.

Once an agent decides to buy cigarettes, they proceed with the actual purchase process. This phase involves several steps:

1. **Cost calculation for each retailer:** Agents calculate the cost of buying cigarette packs (ranging from 1 to 20 packs) from each available retailer. This range aligns with the Tobacco Town model while reflecting a realistic upper limit for typical purchases.

This calculation considers factors such as cigarette prices at the retailer, and costs associated with travel and time. The net cost C_r is given by:

$$C_r(q) = \left(\frac{d_r}{s} + \frac{1}{12}\right) \cdot w + d_r \cdot \frac{P_{fuel}}{E} + q \cdot P_r \tag{4.1}$$

Where:

- $C_r(q)$ is the cost of buying q packs of cigarettes from a retailer
- d_r is the extra distance (Manhattan distance) the agent must travel to the retailer and then to their workplace
- s is the agent's travel speed (km/h)
- $1/12$ represents an average of 5 minutes spent on the transaction (in hours)
- w is the agent's hourly wage (£/h)
- P_{fuel} is the price of fuel (£/L)
- E is the fuel efficiency of the agent's mode of transport (km/L)
- P_r is the price of one pack of cigarettes at the retailer (£)

The formula combines three types of costs (Time, Travel and cigarette cost) to give the total cost for the purchasing decision, allowing agents to compare different retailers and purchase quantities. By including both direct (cigarette price) and indirect (time and travel) costs, the model attempts to capture a more comprehensive view of the economic factors influencing purchasing decisions using a rational agent assumption. While the rational agent assumption is a questionable one, as real-world decision-making often involves non-rational factors, this approach has been validated in the Tobacco Town paper (Luke et al., 2017). It provides a structured framework for modelling complex purchasing behaviours, balancing computational feasibility with a reasonable approximation of economic decision-making processes.

The model adapts to different transportation methods by adjusting the values of time cost and travel cost. For teenagers or those using public transport, walking, or cycling, the travel cost becomes 0, while the time cost increases due to slower travel speeds. This flexibility allows the model to represent various scenarios realistically.

While we acknowledge the limitations of this model, particularly its assumptions about rational decision-making and perfect information, we believe it provides a reasonable approximation of purchasing behaviour for our purposes.

2. **Optimal quantity determination for each retailer:** Agents determine the optimal quantity of cigarettes to purchase from each retailer in the environment. They do this by selecting the quantity that minimises the overall cost per pack(as calculated in step 1). The optimal quantity (number of packs) q_r for each retailer r is calculated as:

$$q_r = \underset{q \in \mathbb{N}^+ \leq 20}{\operatorname{argmin}} \left[\frac{C_r(q)}{\sum_{k=1}^{q} (\delta^{\frac{20.|k-1|+I}{SR}})} \right] \tag{4.2}$$

Where, δ is the discounting rate, which causes agents to devalue cigarette packs that they will smoke later in the future, thus preventing stocking up a lot when their smoking rate is low. I is the number of cigarettes in the agents inventory (in number of sticks rather than packs). SR is the smoking rate per day of the agent.

3. **Cost calculation for optimal quantities:** Agents calculate the total cost of purchasing the optimal quantities of cigarettes from each retailer based on the quantities determined in step 2.

$$C_r(q^*) = \left(\frac{d_r}{s} + \frac{1}{12} \right).w + d_r.\frac{P_{fuel}}{E} + q^*.P_r \tag{4.3}$$

4. **Trembling hand process for retailer selection:** Finally, agents use the "trembling hand process" to decide which retailer to visit and make their purchase. The retailer with the lowest C_r is chosen with a probability $(1-\varepsilon)$, where ε is set at 0.25. The other retailers can be chosen with a probability

$$Prob_r = (1-m)^{rank(r)-1}$$

where *rank*(r) is the rank of retailer r's cost and m is set to 0.5 (chosen from tobacco town paper). The original tobacco town paper, uses a lower value of $\varepsilon = 0.025$. However, we have increased it to 0.25 to offset the high privileges that retailers with lower expected costs have, to make it more realistic. With a very low ε value, agents in the model would almost always choose the lowest-cost retailer, leading to unrealistically concentrated purchasing patterns. By increasing ε to 0.25, we introduce more variability in the decision-making process, potentially mimicking the complexity of real-world consumer behaviour.

However, the realism of this change requires empirical validation. Ideally, the model's outputs should be compared against real-world data on tobacco retailer market shares and consumer purchasing patterns. Due to unavailability of such data in our case, this is a theoretical adjustment based on qualitative understanding of consumer behaviour. Future work should focus on validating this parameter against actual market data.

When agent's age is not in the legal bracket (young-adults)

In situations where an agent's age falls below the legally permissible range for purchasing tobacco products, a strategy is implemented that involves attempting to acquire cigarettes from a retailer. If this attempt is unsuccessful, agents then explore the possibility of purchasing cigarettes from peers or other individuals who are of legal age within their local vicinity.

Attempting to buy from retailers :

1. **Behaviour similar to adult agents**: *Young-adult agents* exhibit behaviour similar to that of *adult agents* when purchasing cigarettes. This behaviour continues up to the point of identifying the retailer and the quantity of cigarette packs through the "trembling hand process."
2. **Additional age verification:** Following the trembling hand process, an extra verification step is implemented, akin to the process followed by *adolescent*

agents (see 4.5.1). This step involves confirming whether the selected retailer is willing to sell cigarettes to individuals who are below the legal age for tobacco purchases.

This additional step ensures that agents below the legal tobacco purchasing age comply with the restrictions and regulations regarding the sale of tobacco products to minors. It adds an extra layer of realism to the simulation, aligning it with real-world legal constraints and purchasing behaviours.

Attempting to buy from social contact : When a *young-adult agent* (Agent X) attempts to purchase cigarettes from a social contact the following steps are employed:

1. **Identification of smoking social contacts:** Agent X identifies all individuals who smoke and are located within a 500-meter radius of both their household and workplace. This radius was chosen to represent a reasonable walking distance for social interactions. One of these potential contacts is randomly chosen (Agent Y).
2. **Checking Agent Y's cigarette Inventory:** If Agent Y's cigarette inventory exceeds Agent Y's daily smoking rate, then agent Y agrees to sell cigarettes to agent X.
3. **Transaction with agent Y:** If Agent Y possesses surplus cigarettes, Agent X can proceed with the purchase. The maximum quantity of cigarettes that Agent X can buy is limited to half of the extra cigarettes that Agent Y has, which go beyond their daily smoking rate. This limitation was introduced to ensure Agent Y retains enough cigarettes for personal use.
4. **Randomised purchase quantity:** Agent X randomly selects a quantity of cigarettes (q_x) to purchase from Agent Y. This quantity falls within the range of 1 to half of the excess cigarettes in Agent Y's inventory.
5. **No purchase if no surplus:** If Agent Y does not possess surplus cigarettes, Agent X abstains from making any additional attempts to purchase cigarettes for the remainder of that day.

It is important to note that, these agents can buy individual cigarettes from their social contacts (not necessarily packs).

Purchasing behaviour for adolescent agents

Current law dictates that adolescent individuals cannot legally purchase cigarettes from retailers. In our model, we aim to simulate the illegal sale of tobacco products to adolescents as well. If the agents are unable to buy cigarettes from retailers (this purchase behaviour is different from that of *young-adults*), they may explore the possibility of purchasing them from peers or other individuals above the minimum legal age in the vicinity, similar to *young-adults*.

Attempting to buy from retailers : This involves several steps:

1. **Retailer selection in vicinity:** Agents attempt to buy cigarettes only from retailers located within a 500-meter radius around their school or household. They then randomly select one retailer from this vicinity as their target for the purchase.
2. **Age verification:** The retailer's decision to sell cigarettes to the agent is determined by a probability value obtained from empirical data (see section 4.7.4). This probability reflects the retailer's willingness or likelihood of selling cigarettes to the agent.
3. **Purchase affordability:** If the retailer agrees to sell cigarettes to the agent, the agent checks whether they can afford to make the purchase based on the selling price set by the retailer. Agents calculate the maximum number of packs of cigarettes they can afford based on the available funds.
4. **Randomised quantity of cigarette packs:** If the agent can afford cigarettes based on the price offered by the retailer, they proceed to buy a certain number of packs of cigarettes (q). The value of q is determined randomly, ranging between 1 and the maximum number of packs of cigarettes they can afford.

Attempting to buy from social contact For *adolescents*, the process of purchasing cigarettes from a social contact follows a similar initial approach to that of *young-adults*. They identify a smoking social contact and check if Agent Y has enough cigarettes. The maximum purchase quantity is determined in the same manner. However, there is an additional step specific to adolescents:

Once the randomised purchase quantity is decided, adolescents only proceed with the purchase if they can afford it using their available pocket money. If they cannot afford the initially determined quantity, they reduce the number of cigarettes they intend to purchase by 1 and repeat this process iteratively until they can afford the maximum number of cigarettes within their pocket money allowance. This assumption provides a clear, consistent rule for agent behaviour, reducing model complexity while still capturing key dynamics of cigarette acquisition.

4.5.2 Smoking initiation

Experimenting smoking

Each *young-adult* and *adolescent agent* who have a smoking status of a never smoker or experimenter smoker, can undergo a daily attempt at experimenting with smoking. This attempt is influenced by a calculated probability that takes into account three factors, (1) Parental Influence: The smoking status of their parents, (2) Peer Influence: The prevalence of smoking among other individuals in their same age group within a 500-meter radius of their household, school/workplace, (3) Addiction Level: The number of lifetime cigarettes the agent has smoked in their lifetime. Section 4.7.3 describes the steps involved in the calculation of this probability in detail.

Becoming a regular smoker

In our model, the definition of a regular smoker follows the guidelines set by the National Center for Health Statistics (US). Specifically, an individual is categorised as a regular smoker if they have consumed a minimum of 100 cigarettes in their lifetime and continue to smoke regularly.

When an experimenter smoker in our model reaches the threshold of having smoked at least 100 cigarettes in their lifetime, they transition to the status of a regular smoker. Furthermore, each regular smoker is assigned a daily smoking rate, which is derived from a distribution of smoking rates based on empirical data (see Section 4.7.4).

4.6 Model initialisation

4.6.1 Retailers

Our model initialises cigarette retailers based on real-world data from Scottish data zones, specifically using the average retailer density per km^2. To achieve this, we first divide our model space into a grid of 1km x 1km squares, achieved by creating 10 divisions along both the x-axis and y-axis, resulting in 100 grid cells. We then iterate through each type of retailer. For each retailer type, we calculate the probability of a retailer's presence in each grid cell as $P(\text{retailer}) = \text{Retailer Density}/100$. This scaling ensures the probability aligns with our 1km x 1km grid structure. Using this probability, we determine whether to place a retailer of that type in each grid cell. This process is repeated for all retailer types, thereby populating the entire grid with a distribution of retailers that reflects the real-world density data. Once the probability condition is satisfied, and a new retailer is allocated to the specific location, we proceed to assign cigarette prices to that retailer. This price assignment is carried out by drawing from a truncated normal distribution, utilising the distribution of cigarette prices available in our data set. This process aligns with the methodology established in the tobacco-town paper (Luke et al., 2017) for consistency and accuracy in our modelling approach.

4.6.2 Schools, workplaces and households

Utilising data obtained from various official Scottish government sources (see Section 4.7.4), we determine the density per square Km of schools and workplaces. We then use this density to calculate the number of schools in each particular environment type (based on the total area) and randomly distribute them throughout the spatial grid. To ensure a more realistic simulation, we have added a condition that if the calculated number of schools is less than 1, we add a minimum of 1 school to the area.

For initiating households, we directly use the average number of households in an environment type from Scottish Survey data. We then employ a similar initiation process as for schools and workplaces, wherein we randomly select locations in the environment grid to assign household locations.

4.6.3 Adolescent agents

1. Proportion-based Allocation: To initialise the *adolescent agent* population, the model first utilises data on the proportion of adolescents (ages 15-18) within the total population of each datazone. Based on these proportions, the number of *adolescent agents* to be generated in the model is determined for each datazone.

2. Assignment of Household and School: Each new *adolescent agent* is randomly assigned a household and a school from the existing virtual households and schools.

3. Age Initialisation: Agents are assigned a random age in days, uniformly sampled between 15 and 18 years.

4. Smoking Status Initialisation: The smoking status of agents (smoker or never smoker) is initialised probabilistically based on age-specific smoking prevalence data from Scotland. We used the prevalence rate available for the age group 15-24, since data specifically for ages 15-18 was unavailable.

   For agents designated as smokers, further classification as experimental or regular smoker is done based on lifetime cigarette consumption. Using empirical evidence on age-specific lifetime cigarette distributions, total consumption is stochastically sampled for each smoking agent from Scottish survey data (see Section 4.7.4). Those exceeding 100 lifetime cigarettes are categorised as regular smokers, with the remainder being experimental smokers by definition.

5. Parental Smoking Status: Parental smoking status is also assigned probabilistically for each agent based on overall smoking prevalence. The probabilities used align with 0, 0.5 or 1 smoking parents (see 4.7.3).

6. Initial pocket money : In our model, all *adolescent agents* within a particular neighbourhood type receive the same amount of pocket money per week. This allowance is calculated as a proportion of the average salary of adults in that specific neighbourhood type.

4.6.4 Young-adult agents

The *young-adult agent* populations are initialised following a similar approach as with adolescent agents, with a few key differences.

First, the initialisation is conducted separately for each transportation type (private vehicle, public transport and others), based on datazone-level proportions of *young adults* and *adults* using each mode.

Other parameters including household, age, and smoking status are assigned analogously as described for adolescents, using age-specific distributions where available.

Instead of being assigned to schools, *young-adults* and *adults* are assigned workplaces randomly from the initialised workplaces, similar to how schools are assigned for *adolescents*.

Additionally, instead of a pocket money allowance, *young-adult* and *adult agents* are assigned an hourly wage drawn from statistical distributions of average wages corresponding to each neighbourhood type (see Section 4.7.4).

4.6.5 Adult agents

We initialise the *adult agents* in a manner very similar to the *young-adult agents* in the model, with one distinction. Since the primary objective of the model is to capture the initiation of smoking, we initialise it with smoker agents only. This approach reduces the model's computational complexity. However, as successive generations of *young-adult agents* transition into *adult agents*, the model will comprise both smoker and non-smoker *adult agents*.

4.7 Overview of simulation steps

A single simulation step corresponds to a day. We differentiate between three distinct age groups of agents: *Adults*, *Young-Adults*, and *Adolescents*, each following specific routines.

4.7.1 Every day

1. *Adults*: Within a day, the activities of *adult agents* are divided into two halves. During the first half, adult smokers assess their cigarette inventory and decide whether to make a purchase.

   If they opt to buy cigarettes, they interact with a retailer and acquire the desired quantity based on their purchase behaviour function. Subsequently, they proceed to their workplaces. In the second half of the day, these adults return home and consume cigarettes according to their smoking rate. Other *adult agents*, such as never smokers or ex-smokers, do not actively contribute to the simulation during this time frame, primarily moving between their residences and workplaces.

2. *Young-Adults*: *Young-adult agents* follow a similar daily routine. They start their day by considering experimentation with cigarettes. Next, they determine their purchase decision (if they are regular smokers) based on their cigarette inventory and proceed to their workplaces. Like *adult agents*, they identify a retailer and purchase cigarettes en route to work, guided by their purchase behaviour function. If their age falls outside of the legal bracket, their behaviour is based on that information. In the second half of the day, *young-adults* return to their homes. Regular smokers within this group adhere to their daily smoking rate while smoking when they have the required number of cigarettes on them (or the maximum available, in case its lower than their smoking rate).

3. *Adolescents*: *Adolescents* also navigate between their households and schools throughout the day. Similar to *young-adult agents*, they start the day by considering experimenting with cigarettes (if they are never-smokers or experimenters). While, the regular smokers attempt to buy cigarettes from the retailers, and if unsuccessful they attempt to buy from other smokers in the vicinity.

A single simulation step represents a day in the real world.

4.7.2 Every 30 steps

At every 30^{th} time step in our simulation, the following events occur:

1. Adolescents ageing into young-adults: *Adolescent agents* whose age has reached or crossed 18 years are transitioned into *young-adult agents*.
2. Young-adults ageing into adults: *Young-adult agents* whose age has reached or crossed 21 years become *adult agents*.
3. Addition of new adolescent agents: New *adolescent agents* are introduced into the simulation based Scottish survey data on number of 16 -year old's. The proportion of the population (16-24) for each environment type was used to calculate this for the simulation setup.
4. Removal of older agents: The agents who are older than 24 years are removed from the simulations.

These periodic events facilitate the dynamic ageing and renewal of agent populations, ensuring that the model represents age-specific behaviours accurately and maintains a realistic demographic structure. We repeat them every at a 30-day interval to decrease the computational costs.

4.7.3 Calibration

To accurately model the transition of experimental agents into regular smokers, we employ indirect calibration. Since the requisite data for this calibration was unavailable for Scotland, we make the assumption that behaviour is similar across different countries. We utilise publicly available data on the number of cigarettes smoked from the U.S. National Youth Tobacco Survey (NYTS) (United States - Centers for Disease Control and Prevention (CDC), 2022), as well as data on parental and peer influences from a study conducted in middle schools in Switzerland and Italy (Scalici and Schulz, 2017). The probability P_{smoke} of an agent experimenting with smoking on a given day is modelled using a logistic regression function incorporating multiple influences:

$$prob = \frac{1}{1 + \exp^{-(\beta_0 + (\beta_1 . x_1) + (\beta_3 . x_3))}}$$

We employ a grid search to optimise the fit between our model's probabilities and the actual probabilities. To understand how good the fit is, we look at the ratio of adolescents who have ever smoked with smoking parents versus those with non-smoking parents. Our model predicts a ratio of 3.5-4.5. This aligns well with empirical findings from various studies, which report this ratio to be between 2 and 5, depending on the specific study. The parameters of the model, calibrated from empirical data, are as follows:

- $\beta_0 = -10.32$, the intercept term.
- $\beta_1 = 1.5$, where x_1 is the parental smoking status (1 if both parents smoke, 0.5 if 1 parent smoker, 0 otherwise). It is worth noting that we do not differentiate between never-smoker parents and parents who have quit smoking, as studies have shown that the offspring of parents who have quit are no more likely to smoke than offspring of parents who have never smoked (Gilman et al., 2009).
- $\beta_2 = 2$, where x_2 is the peer smoking prevalence within 500m radius of home/school (proportion of smoking peers).
- $\beta_3 = 0.6$, where x_3 is the total number of cigarettes the agent has smoked in their lifetime.

This multi-factorial probability calculation aims to capture the complex interplay of parental, peer, and habit influences on an agent's decision to experiment with smoking, making the model behaviour more realistic and data-driven.

4.7.4 Validation

We use the validated Tobacco Town model as the foundation for the development of our own model. The behaviours exhibited by adults in our model are consistent with those in the Tobacco Town model. Our main contribution lies in the addition of behaviours for *adolescents* and *young adults* in scenarios where it is illegal for them to purchase tobacco. To accurately represent this while avoiding over-fitting, we have minimised the number of calibrated variables, opting instead to rely on input validation by using extensive data from Scotland. Details are provided below.

Input validation

In this section, we discuss the details concerning the values of parameters integrated into the new behaviours in the model and their sources. A meticulous input validation process was undertaken to ensure that the model reliably simulates different neighbourhoods in Scotland.

In addition, to substantiate the model's reliability, it was validated against some indicators related to adolescent smoking initiation. The model's projection of the number of adolescents becoming regular smokers, which was found to closely match the general adult smoking prevalence.

Further, our simulation results show that the ratio of adolescents who have ever smoked is 3.5-4.5 times higher if their parents smoke, falling within the empirically observed range of 2-5 (Department of Health and Social Care, 2021; Mays et al., 2014).

The details of the parameters associated with the input validation are described below.

Workplaces

To determine workplace density for model initialisation, data from the Scottish Government's 'Business Sites and Stocks by Urban Rural Classification' was utilised. Specifically, the 2018 dataset was leveraged, providing the latest available business data classified across urban/rural domains (Scottish Government, 2022b).

While datazone-level details were unavailable, totals of registered private business 'sites' by urban/rural classification served as a proxy for workplace density per datazone. Datazone workplace densities were assumed to remain consistent across SIMD quintiles due to limitations in granular data availability. This assumption is a simplification of real-world conditions, as workplace distribution likely varies across areas of different socioeconomic status. However, without more detailed data, this approach allows us to maintain a consistent framework for workplace allocation across the model.

Table 4.2 summarises the parameter values used for different neighbourhoods in the simulation.

Households

For our model, we rely on data from the "Households and Dwellings in Scotland, 2021" survey to determine the number of households (Scottish Government, 2021a). This survey provides valuable information on households in Scotland, including details classified by urban/rural categories and SIMD quintiles. We leverage this data to calculate the number of households for each specific type of neighbourhood within each datazone in our simulation.

Schools

To calculate the density of secondary schools per square kilometre (km^2) in each SIMD and urban/rural classification, we rely on data from the Scottish School Roll and Locations dataset (Scottish Government, 2022a). Given that our model focuses on adolescents, specifically secondary school-aged individuals, we use data exclusively related to secondary schools. The dataset used for this calculation is from the year 2022.

Population estimates by age

To determine the number of agents in each age group for model initialisation, we employ data from the Population Estimates by Scottish Index of Multiple Deprivation

Variable	Most-deprived urban	Least-deprived urban	Most-deprived rural	Least-deprived rural
Area of 1 datazone	0.2438	0.7121	3.4827	21.0050
Workplace density	32.3712	32.3712	0.7909	0.7909
School density	0.0912	0.1324	0.0000	0.0051
Number of households	388.0897	337.0263	347.1957	353.2451

Table 4.2: Summary statistics of environmental agent densities, households, and economic variables used to initialise the model environment, categorised by urban/rural and SIMD quintile classifications. Data sources include various Scottish government statistics from 2018-2022

(SIMD) for the year 2021 (Scottish Government, 2021b). While this data is available at the national level, we utilise it to calculate the proportion of each age group relative to the total population of Scotland.

Once we have established these proportions, we utilise the total number of individuals in each datazone (Scottish Government, 2020) to calculate the number of agents in each age group for every datazone within various neighbourhood types. This approach allows us to distribute agents across different environments while maintaining the age distribution aligned with population data.

In addition, we calculate the proportion of the population who are 15 years old using the provided dataset. This information is used to determine the number of new agents to be introduced into the model.

Table 4.3 indicates the values used in different neighbourhoods in the simulation.

Prevalence

For the initial assignment of smoking statuses to agents as either never smokers or smokers, we rely on prevalence data. Specifically, for individuals aged 15, we use data from the Scottish Schools Adolescent Lifestyle and Substance Use Survey (SALSUS)

Variable	Most-deprived urban	Least-deprived urban	Most-deprived rural	Least-deprived rural
Population	760.4974	805.3308	725.5870	795.6863
15 year old proportion	0.0113	0.0110	0.0113	0.0110
Adolescent proportion	0.0333	0.0327	0.0333	0.0327
Young-adult proportion	0.0305	0.0349	0.0305	0.0349
Adult proportion	0.0363	0.0393	0.0363	0.0393
Pocket money	5.7861	12.6401	5.7861	12.6401

Table 4.3: Population and economic statistics utilised to determine model agent populations and attributes for each neighbourhood type, calculated using 2021 Scottish population estimates and household survey data.

Prevalence	Most-deprived	Least-deprived
Smokers (All ages)	0.5000	0.2200
Never-smoker (15 year old)	0.6557	0.7050
Never-smoker (16-24 year old)	0.5943	0.8164

Table 4.4: Age-specific smoking prevalence estimates used for initial smoking status assignment of agents, obtained from the 2020 SALSUS survey and 2021 Scottish Health Survey. Values shown are for most-deprived and least-deprived neighbourhoods based on SIMD quintiles.

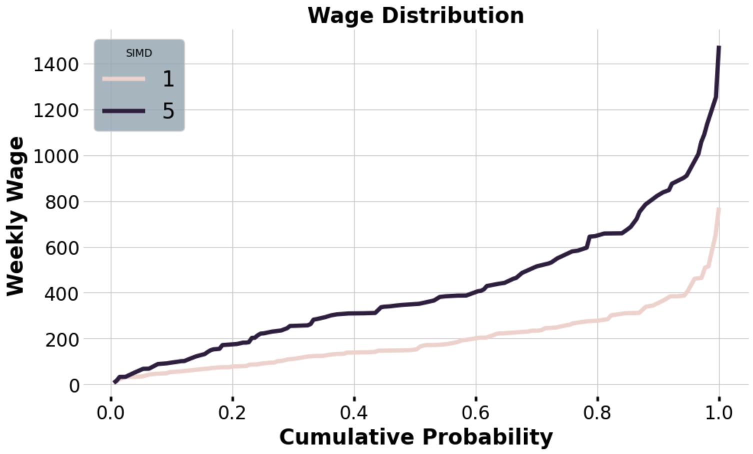


Figure 4.1: Distribution of weekly wages by SIMD quintile used to assign income to agents, based on data from the 2016 Scottish Household Survey.

conducted in 2020 (Ipsos MORI Scotland, 2020). In the case of individuals aged 16-24, we rely on prevalence data from the Scottish Health Survey (SHS) conducted in 2021 (ScotCen Social Research, 2023). Given the limitations of sample size and missing data for specific ages in the SHS, we aggregated the prevalence values for individuals aged 16-24 into a single group.

Additionally, we use the SHS dataset to calculate the prevalence for all adults in each environment as well.

The prevalence values were calculated for each SIMD quintile, as information regarding the urban/rural split was unavailable. As a result, we assumed that the prevalence values were consistent within each SIMD quintile across different urban/rural classifications.

Table 4.4 summarises the values used in the simulations.

Wage distribution

To generate income distributions for the agents in the model, we use the equalised income data from the Scottish Household Survey in 2016 (Ipsos MORI Scotland, 2016a). We convert these income values into hourly wages, which are then used to inform the agent behaviour in the model. Figure 4.1 shows the cumulative wage distribution used in the simulations.

Pocket money - adolescent agents

Average weekly pocket money for adolescents aged 15-18 years was derived from the gohenry report on children's pocket money in the UK (gohenry, 2022). This value was compared to national average adult weekly earnings to determine the proportional ratio of adolescent pocket money to adult income.

Using this ratio, location-specific average adolescent pocket money estimates were calculated for each SIMD quintile and urban/rural classification based on average adult wage data. This method scales adolescent income as a percentage of adult earnings to reflect relative deprivation levels across geographic domains.

Smoking rate distribution

To assign smoking rates to agents, we draw values from the distribution of cigarettes smoked obtained from the Scottish Household Survey in 2016 (Ipsos MORI Scotland, 2016a). This data does not provide granularity by SIMD or urban/rural classification, nor does it differentiate between different age groups. Therefore, we apply the same distribution uniformly across all cases in the model.

Transport

Proportion of transportation modes: We rely on data from the (Ipsos MORI Scotland, 2016b) report to estimate the distribution of transportation preferences among the population. This includes data on the method of travel to work and whether employed adults have the option to use public transport. This information helps us determine the proportion of individuals using cars, buses, or other means of transportation in our model.

Estimated speeds for transportation modes:

1. Cars: The average speed on Local 'A' roads in England across 24 hours in 2021 was estimated to be 38.78 km/h (Department of Transport, United Kingdom, 2022). We use this value as the speed for cars in our model.

Cigarettes per day	Cumulative Probability	Cigarettes per day	Cumulative Probability
1	0.0331	15	0.6907
2	0.0704	16	0.6928
3	0.1024	17	0.6971
4	0.1237	18	0.7051
5	0.2048	19	0.9227
6	0.2251	20	0.9243
7	0.2421	21	0.9477
8	0.2688	22	0.9483
9	0.2741	25	0.9488
10	0.5317	27	0.9824
11	0.5328	29	0.9829
12	0.5605	30	0.9952
13	0.5621	35	0.9973
14	0.5637	40	0.9995
		50	1

Table 4.5: Distribution of number of cigarettes smoked per day for regular smokers in the model, based on data from the 2016 Scottish Household Survey

Variable	Most-deprived	Least-deprived
Car proportion	0.5820	0.7080
Bus proportion	0.1640	0.0770
Other proportion	0.2530	0.2150

Table 4.6: Table presents the proportion of people using various services for commuting to work. The proportions are provided for most-deprived and least-deprived areas based on SIMD quintiles.

2. Buses: We consider average bus speeds for two major cities in Scotland. For Glasgow, the average bus speed is 16 km/h, and for Edinburgh, it is 14.9 km/h (Strathclyde Partnership for Transport, 2022). To simplify, we take the approximate average of these values, which is 15 km/h, as the estimated speed for buses in our model.

3. Other Means of Transportation: We assign a speed of 2.1 km/hr for the "other" category. This lower speed is used to encompass various modes of transportation, including walking, wheelchair use, skateboarding, etc. While this category could also include trains with higher speeds, we assume trains are not available for our model's purposes, as we focus on a small area.

These estimates are used to simulate travel speeds for different modes of transportation within our model, ensuring realistic movement patterns.

Price of cigarettes

The pricing data for cigarettes used in our model is from an unpublished work which used data from The Retail Data Partnership Ltd. This data comprises longitudinal records of tobacco sales spanning the period from 2019 to 2022 and is primarily derived from convenience stores in Scotland. While limited to the convenience retailer type, these comprise approximately 40% of all tobacco retailers and 55-60% of total tobacco sales in the UK (Robinson and Reid, 2017).

The pricing information is available at both the SIMD quintile level and the urban/rural classification levels. However, to ensure robustness in our model and due to limited sample sizes, we opted to use pricing values based solely on the SIMD quintile level. This choice is predicated on the assumption that cigarette prices are likely to exhibit similar trends in both urban and rural settings. The pricing data utilised is from the year 2022. In our analysis, we consider the average price across all cigarette brands rather than differentiating between individual brands.

Retailer Type	Most-deprived Urban	Least-deprived Urban	Most-deprived Rural	Least-deprived Rural	Probability of illegal sale
Off licence	0.5565	0.0543	0.0375	0.0014	9%
Large retailer	0.3436	0.0684	0.0000	0.0042	0
Newsagent	1.0308	0.1085	0.0687	0.0047	18 %
Other	0.4652	0.0456	0.0187	0.0009	0
Petrol station	0.2798	0.0944	0.0312	0.0037	20%
Pub/club	0.5017	0.0901	0.0250	0.0042	0
Private home	0.0365	0.0304	0.0062	0.0019	0
Small retailer	3.5182	0.3299	0.3371	0.0121	24%

Table 4.7: Density of various tobacco retailer types (per km^2) in each environment and the probability of each retailer type selling tobacco products illegally to underage individuals. These values are used to determine retailer placements during model initialisation. The probability of illegal sales remains consistent across different types of neighbourhoods. The density values are derived from tobacco outlet data.

Tobacco retailer density

We utilise data on retailer density per SIMD quintile and urban/rural classification from Scotland to model retailers in our simulations (CRESH Research Group, 2023). Table 4.7 summarises the retailer densities in each of the neighbourhoods that we focus on in our simulations. To maintain consistency in our model, we categorise various types of retailers into eight distinct categories. A study conducted in 2015 assessed each of these retailer categories for their likelihood of illegally selling tobacco products to individuals below the legal purchasing age (Chartered Trading Standards Institute 2015, 2015). We use this information to establish the probability of illegal tobacco product sales within our model.

Using such diverse data sources to inform the model presents several challenges due to inconsistencies in data collection methods, temporal misalignment (with data spanning 2016-2021), and geographical discrepancies (such as using U.S. data when information from Scotland is not available). The model also makes assumptions when applying data, like consistent prevalence across urban/rural areas within SIMD quintiles, which may oversimplify complex patterns. These issues could lead to inaccuracies in the model's predictions, especially for specific subgroups or areas. To mitigate these weaknesses, it's crucial to prioritise recent, Scotland-specific data where possible, conduct sensitivity analyses, clearly document all assumptions and limitations, and regularly update the model with new data.

4.8 Conclusions

In this chapter, we have employed a complex systems approach to develop a spatial, agent-based model that serves as a testbed for a wide variety of tobacco-control policies. The model simulates agents from different age groups, captures their daily mobility patterns to schools and workplaces, and accounts for their transport options. Furthermore, the model replicates interactions among agents as well as between agents and the tobacco retail environment.

Leveraging extensive data from Scotland, we have created diverse simulation neighbourhoods, categorised by deprivation scale and urban-rural classification. The model tracks multiple tobacco-related indicators, such as daily cigarette consumption, total lifetime cigarettes smoked, tobacco retailer sales, and the financial means of individuals across various age groups.

This simulation framework enables us to develop and assess policies targeting different elements within the tobacco landscape, including the tobacco product itself, policies centred on the tobacco user, and those focused on the retail environment. The adaptability of the agent-based model allows for the exploration of various policy combinations and provides insights into their effects on socio-economic indicators.

Chapter 5

Assessing the Impact of Tobacco Control Policies

5.1 Introduction

In the previous chapter, we developed a spatial heterogeneous agent-based framework to model tobacco use behaviour. In this chapter, we utilise this model to analyse the effects of four distinct policy interventions: raising the price of tobacco products, enforcing a ban on cigarette sales around schools, increasing the minimum legal age to purchase tobacco, and a combination of all these interventions. We examine their impact on smoking patterns and also explore how socio-economic differences influence the effectiveness of each policy. Our results suggest that although individual policies have their own benefits, a combined approach is much more efficient in reducing smoking consumption equitably.

5.1.1 Policy landscape

As highlighted in the previous chapter, the tobacco usage and policy landscape is complex, influenced by multiple different factors and their interactions that drive consumption, initiation, and cessation. Social dynamics, such as peer (Harakeh and Vollebergh, 2012; Liu et al., 2017) and parental influences (Gilman et al., 2009), play a crucial role. Spatial considerations, like the accessibility and availability of tobacco products (Cantrell et al., 2015; Marsh et al., 2021; Pokorny et al., 2003; Shortt et al., 2016), also hold significant weight. Moreover, individual demographics, such as age (O'Loughlin et al., 2014), interwoven with socioeconomic factors like financial capacity (Cui et al., 2019) and exposure to smokers and tobacco retailers (Caryl et al., 2020), further complicate the tobacco usage scenario, thereby shaping the policy environment.

Over the years, various policies have been introduced globally, each addressing different aspects of this complex tobacco regulatory landscape. Some of these policies directly target the product, as seen with the rise in the price of tobacco products. The effectiveness of cigarette pricing policies, especially through taxation, has been extensively studied (Gonzalez-Rozada and Montamat, 2019; Organization, 2010; Van Hasselt et al., 2015; Wilkinson et al., 2019). Empirical research suggests that higher prices deter young adults from starting smoking (Jha and Peto, 2014; Lillard et al., 2013). Concurrently, price-based policies have also significantly increased cessation rates. An increase in prices has a direct and immediate effect on the consumption of cigarettes, and the larger the price hike, the greater the reduction in consumption (Jha and Peto, 2014).

Other policies, meanwhile, focus on the spatial exposure and the availability of tobacco products. Strategies that aim to reduce retailer density fall under this category (Glasser and Roberts, 2021). The density of tobacco retail outlets in a particular geographic region has been linked with the prevalence of tobacco use in that area (Luke et al., 2014). Such associations can be attributed to increased accessibility to tobacco products, enhanced visibility which potentially drives consumption, and overall higher exposure (Caryl et al., 2020). To tackle these issues, a variety of interventions have been proposed and tested. These include the banning of tobacco product sales in specific types of stores (Solutions, 2019), prohibiting sales near schools (Henriksen et al., 2008; Mistry et al., 2015; Myers et al., 2015; Ribisl et al., 2017), and reducing the visibility of tobacco products through bans on advertisements at retail locations. In Scotland, for instance, there has been evidence suggesting that reduced visibility through advertisement bans has contributed to lower smoking initiation rates (Haw et al., 2020).

Given the complex nature of tobacco policy, another focus has been on the individual tobacco user itself. Our previous work in Chapters 2 and 3 employed the tobacco contagion model and the friendship paradox, revealing the potential for interventions. Another significant intervention has been the modification of legal regulations concerning the permissible age for tobacco sales. For instance, in the UK, the minimum legal age for purchasing tobacco was raised from 16 to 18. This policy significantly curtailed tobacco access and use among young individuals (Anyanwu et al., 2020). Despite age restrictions, there remains a high risk of young individuals illicitly obtaining tobacco products from retailers (CDC, 2016). However, the overall reduction in consumption is a notable achievement. In recent years, policy researchers have advocated for raising the minimum legal purchase age to 21, drawing parallels to the benefits observed when similar laws are applied to alcohol (Bonnie et al., 2015; Oyston, 2017; Winickoff et al., 2014). The positive impact of increasing the minimum legal age to purchase tobacco is evident in the case of the town of Needham, Massachusetts. They implemented such a policy in 2005 and witnessed a sharp decline in tobacco use prevalence among young individuals (Schneider et al., 2016). By restricting specific age groups from accessing tobacco, these policies play a significant role in reducing both the prevalence and onset of smoking (Fidler and West, 2010)

In addressing the complex nature of tobacco policy, computational modelling offers invaluable insights, allowing us to simulate interactions across individual and societal levels. Our agent-based model, developed in Chapter 4, serves as a tool in this regard. By mimicking real-world scenarios, agent-based models allow us to understand the dynamics of tobacco consumption and the potential impact of different policies across heterogeneous populations.

Using this model, we will simulate the effects of four specific tobacco control policies for a Scottish population. These policies were selected based on their demonstrated effectiveness in empirical studies, as well as through consultations with stakeholders such as Public Health Scotland. The selected policies target diverse aspects: the product itself (by raising the price of tobacco products), the environment (by enforcing a ban on tobacco sales near schools), and the individual (by increasing the minimum purchase age to 21). The fourth policy involves implementing all three of these measures simultaneously. This comprehensive approach enables us to assess not only the individual impact of each policy but also to explore the potential synergies that arise when these policies are combined. Through simulations and detailed analysis of the model results, our objectives are:

1. To assess the impact of each of the policies and its effects on smoking initiation, consumption, and prevalence.
2. To understand the socio-economic implications of these policies.
3. To determine whether a combination of these policies enhances their individual effects by addressing multiple facets of the issue simultaneously.

The simulations aim to provide novel, empirically-grounded insights into the complex systemic implications of different tobacco control policies. By taking a granular agent-based modelling approach, we can evaluate not just aggregate patterns but distributional effects on different age groups across heterogeneous populations and environments.

5.2 Methods

5.2.1 Model

The evaluation of the three policies is carried out using the model introduced in Chapter 4. Our model is a spatial model, we simulate an area of thirty data zones, all of which share the same SIMD deprivation quintile and urban-rural classification. The agents within this model are classified into three age-based groups: *Adolescent agents* aged 15-18, *Young-adult agents* aged 18-21, and *Adult agents* aged 21-24. These agents interact with each other and their surrounding environment, driven by age-specific behaviours detailed in Section 4.5.1.

5.2.2 Policy Design and Stakeholder Consultation

The development of our ABM and the selection of the policy interventions were informed through a stakeholder consultation process aligning with the best practises in participatory modelling (Voinov et al., 2018). We actively worked with researchers from the Centre for Research on Environment, Society and Health (CRESH) at the University of Edinburgh, who brought in expertise in policy analysis and diverse perspectives. This collaborative approach, combined with external stakeholder engagement process ensured that our research remained both scientifically and practically relevant.

Our project benefited from two key groups of contributors. (1) Internal Collaboration: Researchers from CRESH were an integral part of the project fro the start. They provided complimentary expertise, participated in all stakeholder meetings and offered insights in to the policy landscape. Additionally, they granted us access to data (see 4.7.4), further improving our research. (2) External stakeholders: Our primary external stakeholder was Public Health Scotland, with whom we engaged at various stages of the project. Additionally, we also presented our work to a broader audience including Scottish Government officials, Action on Smoking and Health (ASH) Scotland and other researchers at a later workshop.

Consultation occurred at various stages:

1. Project initiation: Discussion of initial ideas and project scope.
2. Mid-modelling and groundwork for interventions: We presented the initial framework of the model and discussed potential increases in model capabilities. We expanded our model to include age-related behaviour, and adolescent behaviour based on potential interventions of interest.
3. Model Completion: Presentation of the developed ABM, including a detailed explanation of the model's assumptions and limitations.
4. Intervention discussion: Interventions were chosen based on multiple factors: practicality, need for intervention, model capabilities, and examples from literature and international practices.

5. Result discussion: Analysis and interpretation of interventions.

The repetitive process ensured that stakeholder input was incorporated at every stage of the project.

Our stakeholders, while not technical coders, had a strong awareness of modelling approaches and had worked with other modellers in the past. This experience allowed for more productive discussions around modelling capabilities and limitations. To ensure clear communication, we engaged the stakeholders with technical presentations, simplified, yet accurate written documentation of the model and presentations tailored for broader audiences at a workshop.

The stakeholders provided crucial insights into intervention practicality as well as directions in terms of interventions. For instance, discussions with the stakeholder led us to consider the policy of increasing the minimum purchasing age from 18 to 21, inspired by New Zealand's age-related interventions. Additionally, their involvement significantly influenced our research direction, particularly in emphasising the importance of understanding socioeconomic effects of interventions.

While stakeholder input was invaluable for understanding practical implications and policy relevance, we maintained full control over the technical modelling process. All model components and assumptions were grounded in scientific literature. Stakeholders encouraged the development of an easily understood yet scientifically solid methodology, aligning with the approaches advocated in literature (Badham et al., 2018) for ABM in policy contexts.

5.2.3 Implementing policies

P1: Raising the price of tobacco products

The first policy assessed focuses on increasing the price of tobacco products throughout the modelled area. In March 2023, the UK Government implemented a rise in tobacco product duty, setting it at 2% above the prevailing Retail Prices Index (RPI) (HMRC, The govenrment of UK, 2023). Based on recent UK government data, the RPI just prior to this adjustment was 12.6%. We use this as a reference point, and in our simulations, we raise the price by 14.6%.

Existing literature points out that cigarette consumption patterns change due to an increase in price. The price elasticity of tobacco consumption is an economic indicator used to measure this change in tobacco consumption resulting from price alterations. It gauges the extent to which the quantity demanded of tobacco alters in response to a price change. We use values from the HMRC, Government of UK report on cigarette consumption to deduce the values for price elasticity in our model for each socio-economic area (Czubek and Johal, 2010). The report indicates that the value has a confidence interval between -1.07 and -1.35. Additionally, existing literature suggests that the impact of price rises is more pronounced on individuals from more deprived areas (Wilkinson et al., 2019). Since we employ the 5-step categorisation based on SIMD for deprivation, we distribute values from within the confidence interval equally across each SIMD quintile: SIMD1 (most deprived): -1.35, SIMD2: -1.28, SIMD3: -1.21, SIMD4: -1.14, SIMD5 (least deprived): -1.07.

When we initiate each agent and if policy P1 is in place, the smoking rate of each individual is adjusted to account for the price elasticity in their respective socio-economic area.

In our model, agents have a price sensitivity function based on their socio-economic background, which affects their consumption of cigarettes.

P2: Ban of sale of tobacco around schools

To implement the second policy, P2, which bans the sale of tobacco products around schools, we follow a two-step process:

1. We utilise the same method as when no policies are applied to initialise the model with retailers. These retailers are randomly distributed within the simulated neighbourhood based on their density from empirical data in a respective datazone.

2. If policy P2 is in effect, all retailers within a 500-metre radius of each school are removed or deactivated.

P3: Raising the minimum legal age to purchase tobacco to 21 (T21)

Policy P3, which raises the minimum legal age for purchasing tobacco products from 18 to 21, primarily targets the *young-adult agents* in our simulations and influences their behaviour. When the minimum legal age is increased to 21, the likelihood of a sale of tobacco from a retailer to a *young-adult agent* becomes similar to that for *adolescent agents*. The key difference, however, is that *young-adult agents* continue to receive a salary rather than pocket money. Their purchasing behaviour is further elaborated upon in Section 4.5.1.

We also account for the probability of illegal tobacco product sales by various retailers. We utilise data on illegal sales to *adolescent agents* and apply these probabilities to *young-adult agents* as well.

5.2.4 Experimental setting

In this chapter for each policy, we run simulations for every combination of urban-rural and SIMD quintile-based areas. We focus on the results from the four extreme cases (most deprived urban and rural, least deprived urban and rural) to gain a deeper understanding of the dynamics. We run the simulations for a period of 10 years. Each simulation was carried out across a total of 30 data zones. This number was chosen to balance computational demands with the need for a population size in which each age group has a sufficient number of agents to maintain realism. We repeat each simulation 100 times to minimise the effect of randomness.

The model was developed in Python using the MESA framework (Kazil et al., 2020).

Indicators used for analysis

To monitor the impact of each policy, we analyse the average values of certain statistics over the final six months of the 10-year-long simulations. Our key areas of focus are:

1. Prevalence: This captures both regular smokers and experimenters (those who have tried at least one cigarette) among *adolescents* and *young-adult agents*.

2. Total Daily Consumption: Representing the cumulative number of cigarettes smoked daily by all agents in an age group.
3. Lifetime Consumption: Indicating the average number of cigarettes an agent in each age group has smoked over their lifetime.
4. Retailer Sales: We assess the number of cigarette packs sold by retailers to each age group. It is important to note that when the agent's age group is below the minimum legal age limit for purchasing tobacco products, this metric represents the illicit sale of cigarettes.
5. Expenditure on Tobacco: Lastly, we analyse the average amount spent on tobacco per pack by both *young-adult* and *adult agents* within the model. This amount takes into account the cost of travel, time, and the price of cigarettes. We also analysed these expenses as a fraction of weekly wages.

We evaluate the relative changes in each of the aforementioned indicators by comparing scenarios where each of the policies are applied against those where it is not. Additionally, we perform an independent t-test ($\alpha = 0.05$) to ascertain that the distribution of the means of these indicators, across various simulation runs, differs significantly from scenarios without the policy implementation or not.

5.3 Results

For each of the policies tested, we evaluate their impact using previously described indicators, recorded at every time-step throughout the simulations. We focus our analysis on the final six months of a decade-long simulation period, highlighting the changes observed when a policy is implemented compared to when it is not. The data is stratified by population categories—adolescents, young-adults, and adults—and further subdivided by neighbourhood types: Most Deprived Rural, Least Deprived Rural, Most Deprived Urban, and Least Deprived Urban.

5.3.1 P1: Price increase

Table 5.1 presents the effects of Policy P1, which focused on raising tobacco prices, on multiple tobacco-associated indicators monitored in the simulations. The results are summarised below.

1. **Total daily cigarettes smoked**:

   When tobacco prices rise, a corresponding decrease can be seen through the effects of price elasticity. The magnitude of this decline varies across socio-economic areas, influencing the total number of cigarettes consumed by smoker agents daily in a particular area. Notably, the price elasticity effect is more pronounced in deprived regions, suggesting a more significant drop in consumption in these areas.

   We observed a decline in the total daily cigarettes smoked by all *adolescent agents* across majority of the socio-economic areas. Significant reductions, with a significance level set at $\alpha = 0.05$, of -10.50%, -8.75%, and -7.56% were observed in the most deprived urban, most deprived rural, and least deprived urban areas, respectively. Although the least deprived rural areas showed a higher reduction of -11.33%, this decline was not statistically significant. This anomaly suggests the presence of outlier values in the simulation results and high variance in the distributions possibly due to adolescents smoking lower number of cigarettes, and them being able to buy individual cigarettes from other smokers.

   Both the *young-adult* and *adult agents* exhibited significant ($\alpha = 0.01$) relative reductions in daily cigarette usage across all regions. Among the *young-adult agents*, the most substantial decrease was in the most deprived rural areas at -17.40%. For *adult agents*, the largest relative reduction was noted in the most deprived rural areas, with a decline of -30.47%.

2. **Lifetime cigarettes smoked**:

   When examining the total cigarettes smoked by all agents throughout their lifetimes, a significant reduction in cigarette consumption is observed across all socioeconomic areas. This trend aligns with the findings from the total daily cigarettes smoked.

Following the trends observed in the case of total daily cigarettes smoked, the *adolescent agents*, the policy P1 resulted in consistent declines among the *adolescent agents* across all regions. The most substantial relative decline occurred in the most deprived urban areas at -11.34%. Sequentially, the most deprived rural areas followed with a reduction of -10.03%, then the least deprived urban areas at -9.21%. The least deprived rural areas also exhibited a decrease of -7.67%. However, this change similar to the case of total daily cigarettes smoked, was not significant in its relative magnitude.

The *young-adults agents* experienced significant reductions across all areas, with the largest relative decrease in the most deprived rural region at -16.79%. Meanwhile, *adult agents* had relative reductions significant across areas, the most substantial being in the most deprived rural region at -21.34%.

3. **Retailer sales**:

   In addition to monitoring consumption, we also observed the sales of tobacco products to each age group. In alignment with consumption trends, the sale of tobacco products by retailers also saw a significant decrease, a trend that was consistent across most of the socio-economic areas.

   The sale of tobacco illegally to *adolescent agents* experienced the most pronounced drop in the most deprived rural regions at -15.86%. The smallest decline occurred in the least deprived urban areas where individuals had higher disposable incomes and where the implications of price elasticity were less pronounced. The least deprived rural regions also reported a decrease of 15.70%. However, this reduction was not statistically significant at the $\alpha = 0.05$ level.

   For the *young-adults agents*, the policy implications were similar, with the most significant decline being -17.07% in the most deprived rural areas, followed by a -12.92% reduction in the least deprived urban areas. Similarly, *adult agents* followed a consistent trend, with the most substantial relative decrease observed in the most deprived rural areas at -25.99%.

4. **Money spent on tobacco**:

   In addition to evaluating the policy's effectiveness in reducing cigarette consumption, we also evaluated its financial impact on individuals from different socio-economic backgrounds.

   The *young-adults agents* saw a relative increase in the amount spent on tobacco across all regions. The most significant increase was observed in the most deprived rural areas, with a hike of 13.60%. However, considering the varied economic backgrounds of individuals, the increase was not consistent when evaluated as a proportion of their weekly salary. For instance, while the *young-adults agents* in the most deprived rural area allocated 2.8% of their weekly salary to a single tobacco purchase, their counterparts in the least deprived urban area expended only 1.32% of their weekly earnings on a single purchase. The agents in the least deprived rural and most deprived urban areas allocated approximately 2.39% and 2.34% of their weekly wages, respectively. It is important to note that rural areas typically have fewer retailers, leading to increased tobacco purchase costs due to higher travel and time expenses.

   Similar to the *young-adults agents*, the *adult agents* also experienced a relative increase in their tobacco expenditure across all regions. The most deprived urban area witnessed the most considerable relative surge, at 12.76%. Nevertheless, this increase was not consistent when assessed as a fraction of their weekly income. Much like the *young-adults agents*, the adults displayed a similar pattern of financial disparity in terms of their spending on tobacco. The resemblance in these numbers are due to both age groups exhibiting analogous behaviours and sharing a consistent wage distribution.

Indicator	Agent	Most Deprived Rural	Least Deprived Rural	Most Deprived Urban	Least Deprived Urban
Daily cigarettes smoked	Adolescents	-8.75 %	-11.33 %	-10.50 %	-7.56 %
	Young-adults	-17.40 %	-15.56 %	-16.35 %	-13.82 %
	Adults	-30.47 %	-27.67 %	-28.54 %	-26.42 %
Lifetime cigarettes smoked	Adolescents	-10.03%	-7.67 %	-11.34 %	-9.21 %
	Young-adults	-16.79 %	-14.49 %	-16.11 %	-12.40 %
	Adults	-21.34 %	-20.56 %	-19.19 %	-18.71 %
Retailer sales	Adolescents	-15.86 %	-15.70 %	-15.68 %	-9.75 %
	Young-adults	-17.07 %	-15.29 %	-15.93 %	-12.92 %
	Adults	-25.99 %	-22.27 %	-25.02 %	-20.28 %
Money spent on tobacco	Young-adults	13.60 %	8.35 %	12.16 %	11.75 %
	Adults	10.99 %	7.59 %	12.76 %	10.92 %

Table 5.1: Effects of Policy P1: Raising Tobacco Prices on Tobacco-Related Indicators. This table illustrates the relative changes in various indicators during the last 6 months of a 10-year simulation period comparing outcomes when Policy P1, focused on increasing tobacco product prices, was applied versus when it was not. The data is presented for different population groups and across various neighbourhood types, with all values expressed as percentage changes. Each simulation was repeated 100 times to account for variability.

5.3.2 P2: Ban on sales of tobacco

Building upon the aforementioned findings related to Policy P1, the Table 5.2 evaluates the effects of Policy P2, which introduces a ban on cigarette sales within 500 meters of schools.

1. **Daily cigarettes and lifetime cigarettes smoked**: When we examine the two indicators for cigarette consumption daily cigarettes smoked and lifetime cigarettes smoked and compare them between scenarios both with and without the application of policy P2, relative change were observed. However, these changes failed to achieve statistical significance. This observation reinforces the argument that the policy does not yield a significant reduction in daily or lifetime cigarette consumption across different age groups.

2. **Retailer sales**: Contrary to the observations regarding daily and lifetime cigarette consumption, policy P2 had a statistically significant impact (at $\alpha = 0.05$) on the illegal sale of tobacco to *adolescents agents* in certain regions. Specifically, in the most deprived rural and least deprived urban areas, the policy's effects were evident. In these regions, policy P2 resulted in a -7.66% reduction in illegal sales in the most deprived rural areas, and a decline of -7.20% in the least deprived urban areas.

   However, when evaluating the effects of policy P2 on the *young-adult agents* and *adult agents*, there was no noticeable difference in terms of illegal tobacco sales. This suggests that the policy's influence is age-specific and most impactful on the adolescent demographic within certain socio-economic settings.

3. **Money spent on tobacco**:

   The implementation of policy P2 did not significantly influence the spending behaviours of *young-adult agents* and *adult agents* with respect to their tobacco purchases. When comparing the expenditures on tobacco in scenarios with and without the policy applied, there were no notable disparities due to the policy in the financial behaviours of these two age groups.

	Agent	Most Deprived Rural	Least Deprived Rural	Most Deprived Urban	Least Deprived Urban
Daily cigarettes smoked	Adolescents	-7.62 %	4.80 %	0.57 %	-3.91 %
	Young-adults	-1.83 %	-3.11 %	0.58 %	2.03 %
	Adults	-3.19 %	0.61 %	-2.49 %	-1.32 %
Lifetime cigarettes smoked	Adolescents	-5.39 %	3.17 %	1.43 %	-7.48 %
	Young-adults	-1.28 %	-2.95 %	-0.28 %	1.31 %
	Adults	-4.64 %	0.09 %	-2.03 %	-1.72 %
Retailer sales	Adolescents	-7.66 %	-0.10 %	0.15 %	-7.20 %
	Young-adults	-1.34 %	-1.69 %	-0.13 %	1.19 %
	Adults	-0.31 %	-0.86 %	0.64 %	1.69 %
Money spent on tobacco	Young-adults	0.75 %	-0.04 %	0.38 %	0.92 %
	Adults	-0.42 %	-0.82 %	0.42 %	1.50 %

Table 5.2: Effects of Policy P2: Ban of sale of cigarettes 500 metres around schools on Tobacco-Related Indicators. This table illustrates the relative changes in various indicators during the last 6 months of a 10-year simulation period comparing outcomes when Policy P2, focused on increasing tobacco product prices, was applied versus when it was not. The data is presented for different population groups and across various regions, with all values expressed as percentage changes. Each simulation was repeated 100 times to account for variability

	Agent	Most Deprived Rural	Least Deprived Rural	Most Deprived Urban	Least Deprived Urban
Daily cigarettes smoked	Adolescents	-4.29 %	-10.23 %	-4.26 %	-5.44 %
	Young-adults	-75.37 %	-84.30 %	-79.35 %	-80.90 %
	Adults	-0.88 %	0.93 %	1.57 %	0.76 %
Lifetime cigarettes smoked	Adolescents	-5.30 %	-4.76 %	-1.23 %	-9.45 %
	Young-adults	-65.18 %	-80.89 %	-65.44 %	-60.86 %
	Adults	-44.24 %	-52.21 %	-44.47 %	-46.29 %
Retailer sales	Adolescents	-3.31 %	1.35 %	3.24 %	2.32 %
	Young-adults	-75.68 %	-84.01 %	-79.24 %	-81.75 %
	Adults	-0.94 %	0.24 %	2.34 %	0.64 %
Money spent on tobacco	Young-adults	0.54 %	3.40 %	-2.77 %	-0.68 %
	Adults	-0.12 %	-6.33 %	0.03 %	-0.73 %

Table 5.3: Effects of Policy P3: Raising the minimum legal age to purchase tobacco, on Tobacco-Related Indicators. This table illustrates the relative changes in various indicators during the last 6 months of a 10-year simulation period comparing outcomes when Policy P1, focused on increasing tobacco product prices, was applied versus when it was not. The data is presented for different population groups and across various regions, with all values expressed as percentage changes. Each simulation was repeated 100 times to account for variability

Policy P2 where retailers are banned from selling tobacco products around schools, does not achieve the desired effect as seen with Policy P1. This approach essentially translates to a localised reduction in retailer density. Consistent with existing research on the subject, the impact of retailer density reduction policies only materialises when there is a substantial decrease in density. The modest reduction in retailer density brought about by banning sales around schools is not sufficient to create significant change. Hence, consistent with existing literature, Policy P2 does not bring significant effects on most indicators, with the exception of a decline in illegal sales of tobacco to adolescents.

5.3.3 P3: Raising the minimum legal age to purchase tobacco

Building on the findings from the previous policies, we analyse the implications of Policy P3, in which we raise the legal purchasing age of tobacco to from 18 to 21 (see Table 5.3).

1. **Daily cigarettes smoked**: The results suggest that while there is a trend of reduced cigarette consumption among *adolescent agents*, the change is not statistically significant at the $\alpha = 0.05$ level.

   The *young-adult agents* exhibited significant reductions in daily cigarette consumption, with decreases ranging from 75% to 85% across regions. The most pronounced decline was in the least deprived rural areas at -84.30%.

   In contrast, *adult agents* does not have a significant reduction in the daily cigarettes smoked.

2. **Lifetime cigarettes smoked**:

   While *adolescent agents* exhibited a decline in lifetime cigarette consumption, this change was not statistically significant.

   Young-adults notably witnessed significant reductions, with the least deprived rural area having a massive decline of -80.89%.

   For *adults*, the most significant reductions were in the least deprived rural (-52.21%) and urban (-46.29%) regions. Given that this metric captures the cumulative effects of daily smoking, policy P3 had a statistically significant impact across all socio-economic areas.

3. **Retailer sales**:

   Sales to *young-adults agents* from retailers saw notable reductions, aligning with the trends in daily cigarette consumption, with decreases spanning from 75% to 85% across various regions.

   Both *adolescent agents* and *adult agents* exhibited no significant shifts in retailer sales.

4. **Money spent on tobacco**:

   The implementation of policy P3 did not significantly influence the spending behaviours of *young-adult agents* and *adult agents*.

In conclusion, the introduction of policy P3 achieved its intended effect on *young-adult agents*, leading to a pronounced decline in their daily tobacco consumption. This reduction, in turn, influenced a decrease in the lifetime tobacco use among both *young-adult* and *adult agents*.

For *adult agents*, the lack of change in their daily cigarette consumption can be traced back to modelling decisions. Specifically, when a *young-adult agent* is faced with limited cigarette access, their inherent smoking rate remains constant. As a result, they try to smoke based on their predetermined rate when cigarettes are available. Consequently, as they mature into *adult-agents* and cigarettes become more accessible, they revert to their original smoking patterns.

Given that Policy P3 is specifically aimed at agents aged 18-21, it logically follows that its impact on *adolescent agents* would be indirect and minimal, a trend evident in the observed results.

5.3.4 P4: Combination of polices

In the following section we evaluate the impact of Policy P4, which combines all three previously explored policies – raising cigarette prices, enforcing a ban on tobacco sales near schools, and raising the minimum legal age to purchase tobacco to 21. The results from the simulations are presented in Table 5.4.

1. **Daily cigarettes smoked**:

   In *adolescent agents* there was a sharp declines in daily cigarette consumption across all regions, with the steepest drop seen in the least deprived urban areas at -22.43% aligning with the results from P1.

   Both *young-adults* and *adult agents* there were very high reductions in daily cigarette consumption, aligning with the previous patterns noted in policies P3 and P1. The least deprived rural area again observed the most significant decline in young-adults at -84.60%.

2. **Lifetime cigarettes smoked**:

	Agent	Most Deprived Rural	Least Deprived Rural	Most Deprived Urban	Least Deprived Urban
Daily cigarettes smoked	Adolescents	-16.90 %	-18.59 %	-12.61 %	-22.43 %
	Young-adults	-75.85 %	-84.60 %	-79.71 %	-81.09 %
	Adults	-31.01 %	-27.50 %	-31.52 %	-26.65 %
Lifetime cigarettes smoked	Adolescents	-16.76 %	-13.45 %	-11.64 %	-22.29 %
	Young-adults	-66.83 %	-81.38 %	-67.81 %	-65.56 %
	Adults	-56.49 %	-63.02 %	-58.82 %	-57.12 %
Retailer sales	Adolescents	-19.35 %	-19.68 %	-14.09 %	-19.86 %
	Young-adults	-76.49 %	-85.19 %	-80.67 %	-81.43 %
	Adults	-26.39 %	-21.42 %	-24.35 %	-19.51 %
Money spent on tobacco	Young-adults	11.72 %	5.85 %	10.03 %	11.32 %
	Adults	11.89 %	9.28 %	13.25 %	13.50 %

Table 5.4: Outcomes of Policy P4: Cumulative Impact of P1, P2, and P3 on Tobacco-Related Indicators. This table provides insights into the combined effects of the three policies over the last 6 months of a 10-year simulation period. Comparisons are drawn between outcomes when the comprehensive Policy P4 was applied against scenarios without it. Presented data spans various demographic groups and regional classifications, with values depicted as percentage changes. Each simulation was reiterated 100 times to factor in variability.

Similar to daily consumption, *adolescent agents* across all regions had considerable reductions in lifetime cigarettes smoked. The steepest decline occurred in the least deprived urban region at -22.29%.

Young-adults and *adult agents* followed a similar trend, with high reductions across all areas.

3. **Retailer sales**:

   Adolescent and *adult agents* showed consistent declines in retailer sales across regions. For *adolescents*, the least deprived urban region marked the highest drop at -19.86%.

   Young-adult agents, similar to their consumption patterns, demonstrated significant declines across all regions.

4. **Money spent on tobacco**:

   Both *young-adults* and *adult agents* displayed an upward trend, signifying an increase in the proportion of their wages spent on tobacco. This trend was especially prominent in the most deprived urban regions, with values reaching 13.25% for adults.

Upon evaluating the impact of Policy P4, it becomes evident that the advantages observed in earlier policies are carried over. There is a significant decrease in both daily cigarette consumption and lifetime cigarettes smoked among adolescents. This reduction becomes even more pronounced when transitioning from *adolescent agents* to *young-adult agents*. Moreover, *adult agents* display considerably larger reductions compared to the impacts of previous policies.

An assessment of cumulative sales data of retailers highlights a consistent trend. Given the substantial influence of P3, a high decline in tobacco sales is also observed, especially among the *young-adult agents*.

5.4 Discussion

In this chapter we have used our agent-based model to examine the impact of three distinct type of tobacco policies both in isolation and combination. Policy P1 tackled product affordability by raising prices; Policy P2 shifted focused spatial availability, implementing a ban on retailers near schools; and Policy P3 specifically targeted the tobacco user demographic via age restrictions. The model captures the complex interactions between agents, behaviour of different age groups including social influence, purchasing behaviour and mobility. Through simulations we test the impact of each of these policies on equity using real-world data on socio-economic deprivation and urban rural classification. The variety of policies examined highlights the adaptability of agent-based models and emphasises the necessity for a complex systems modelling approach in tobacco control research.

Impact of policies A direct effect on daily cigarette consumption and lifetime cigarette smoked is observed when tobacco prices are increased. This effect was consistent across all age groups, and could be seen from the retailer sales results as well. Consistent to empirical studies, our findings suggest that impact is higher in more deprived areas where the average salary is lower (Wilkinson et al., 2019). A similar trend was observed when when we looked at impact based on urban-rural classifications. Nevertheless, it is crucial to note that while this policy effectively reduces smoking, it simultaneously amplifies financial burdens on individuals, particularly in more deprived areas, a trend empirically observed in studies conducted outside of Scotland as well (Blakely and Gartner, 2019; Guillaumier et al., 2015; Hoek and Smith, 2016).

The policy that raised the minimum legal age for cigarette purchases from 18 to 21 (P3) showcased a high reduction in cigarette consumption. Since this policy was mainly focused on the age group consisting of the *young-adult agents* in our model, the impact was seen the most in the same age group. Despite there being differences in the relative drop between different socioeconomic areas, the decline was very high in all of them.

In contrast to P1 and P3, the evaluation of policy P2 across diverse metrics revealed its limited efficiency in influencing the consumption patterns of agents from different age groups and socio-economic backgrounds. While the policy did not substantially alter daily or lifetime cigarette consumption nor did it increase the financial burden on individuals non-equitably. However, the impact of the policy P2 becomes evident in specific socio-economic areas. In the most deprived rural and least deprived urban areas, there was a significant reduction in the illegal sales of tobacco to *adolescent agents*. This suggests that while the policy P2 might not have a broad impact on consumption or spending behaviours, it has potential in addressing the issue of illegal sales in certain locales.

A key aspect to consider with the implementation of Policy P2 is the potential for a normalisation effect. By reducing the availability of tobacco retailers in proximity to adolescents, the policy might limit their exposure to the sale and advertising of cigarettes to adolescents. Over time, this decreased exposure can contribute to reshaping societal norms and perceptions around smoking, making it less of a commonplace or normalised behaviour among the younger population. This normalisation effect can have a long-term impact on reducing smoking initiation rates among adolescents, as they might no longer view smoking as a standard or prevalent practice among their peers. It also has the potential to curb the temptation of smoking that comes from its perceived ubiquity in society. However, our current model cannot capture this dynamic fully. It remains a challenge to accurately quantify the influence of decreased visibility of tobacco retailers on the behaviours and perceptions of adolescents. Nevertheless, it is an avenue worth exploring in the future as it offers insights into more subtle, yet highly impactful effects of policies like P2.

From our simulation results, we can see that policy P4 successfully integrates the beneficial outcomes of the individual policies while addressing some of their limitations. P4 enhances the impact seen in Policies P1 (price increase) and P2 (sales ban near schools) on adolescent agents across all socio-economic regions. For instance, while P1 reduced the lifetime cigarettes smoked by adolescents by 9.21% and P2 by 7.48%, P4 achieved a 22.29% reduction, surpassing the sum of their individual effects.

P4 maintains the strong impact of Policy P3 (raising the minimum age) on young-adult agents across all socio-economic areas. For eg., the reduction in the lifetime cigarettes smoked by young adult in most deprived rural areas under P4 (66.83%) is comparable to P3 (65.18%), ensuring that the significant benefits of age restrictions are not diluted in the combined approach.

Although impact of raising the prices as a financial burden persists, it is slightly alleviated compared to the effects of Policy P1 in isolation. The inequality in expenditure, observed as a fraction of weekly wages, persists; however, its magnitude does not increase.

Efforts to bridge the gap between rural and urban areas have shown progress. Nevertheless, it is evident that the most deprived urban regions still lag behind, recording the smallest declines across the simulated socio-economic areas.

Raising the minimum legal age for tobacco purchases can inadvertently promote illegal sales. While our model demonstrates that, despite an increase in illegal sales, overall tobacco use experiences a significant reduction. In our simulation, due to data constraints, we assume that the likelihood of illegal sales to *young adult agents* is equivalent to that for *adolescent agents*. However, this might be a conservative estimate. The societal norms, which often lean towards greater individual freedom for young adults, can potentially lead retailers to sell tobacco products illicitly to young adults at a higher rate compared to adolescents. Ensuring compliance therefore remains a crucial facet of any policy implementation.

Recommendations for policymakers Our findings further emphasise the efficacy of a multi-faceted policy approach (Flor et al., 2021; The Lancet Public Health, 2019), particularly when combining elements from Policies P1, P2, and P3 into Policy P4. Policymakers should consider:

1. Prioritising combined strategies over individual policies, as they show a more substantial and widespread impact across different demographic and socio-economic groups.

2. Monitoring and adapting price-based interventions, given their potential adverse financial consequences on certain socio-economic segments.
3. Investing in awareness campaigns alongside implementing policies, especially in the most deprived urban areas, to ensure better compliance and understanding of the regulations.

Limitations Our agent-based model effectively evaluates several cigarette consumption and sales indicators, however it can be improved by addressing some inherent limitations from modelling choices and data constraints.

One such limitation is on the prevalence estimates. Although our model adeptly captures the determinants leading to smoking initiation, it lacks a mechanism for quitting behaviour. Given that our primary objective was to elevate the model standards by integrating a realistic depiction of social initiation factors, introducing a quitting behaviour based on reduced consumption was not the focus of this research. The integration of an empirically grounded cessation behaviour would enhance the model and lead to accurate prevalence simulations and more granular insights into policy effects. We identify this integration as a priority for future research.

Our model heavily relies on a cost function for cigarette purchases. Although based from current existing literature, this function is a simplified representation of the complex economic decisions individuals make. The equation-based approach, while allowing computational efficiency, may not fully capture the real-world purchasing behaviours. The quality of this representation could be improved by incorporating more nuanced economic models and considering individual variations in purchasing behaviour.

We have used simplified representations of initial smoking rates in our model with agents having the same distribution of cigarettes smoked per day across all demographic groups. Similarly, our model assumes a fixed ratio of pocket money to adult income and that young adult agents work and have a salary (part-time or full-time). These simplification, while necessary due to data limitations, do not fully reflect the heterogeneity present in real populations. This potentially potentially impacts the models ability to accurately represent subgroups within the initialised population. Further iterations could incorporate more granular demographic data and smoking rates to improve this aspect of the model.

Our model also assumes that the probability of illegal sales to young adults is equivalent to that for adolescents. While this is likely a conservative estimate, it's an oversimplification. Additionally our model does not account for how retailers might adapt their behaviours in response to policy changes over time, which could affect long term impact of certain policies.

The initiation process in our model, while incorporating factors like peer influence and parental smoking status, is ultimately based on an equation with best-fit parameter values. While these values are calibrated to available data, they may not perfectly represent true values in all contexts, potentially affecting the accuracy of our initiation predictions.

Lastly, our model strongly relies on spatial dynamics to simulate social contagion amongst peers. A more refined strategy would involve the use of social network-based interactions that could more authentically capture the spread of smoking behaviour. However, adopting this methodology will require access to more comprehensive data than what is currently available. Implementing such a model would require considerations for birth and death dynamics, as well as information on interactions both within and across different age groups. Moreover, integrating this social network aspect with spatial elements would demand a sophisticated multi-layered network methodology. While we recognise the potential of such an approach in improving the precision of modelling tobacco usage behaviours, we reserve its incorporation for subsequent research.

Despite these limitations, our model provides valuable insights into policy impacts and serves as a foundation for future refinements in tobacco use modelling. Future work should focus on addressing these limitations to enhance the model's precision and applicability.

5.5 Conclusion

Despite global declines in tobacco use, its persistent challenge highlights the need for ongoing and innovative policy interventions. Our agent-based model offers a detailed analysis of tobacco use patterns in a representative Scottish population, with an emphasis on the impact of distinct tobacco control policies in areas segmented by socioeconomic deprivation and urban-rural classification.

Based on our simulations, the policies yielded the following impacts:

1. Price Rise (P1): Displayed a significant effect across all demographics, curbing smoking initiation and encouraging cessation. However, it introduced a financial challenge, particularly in more deprived areas.
2. Sales Ban Near Schools (P2): Primarily impacted illegal tobacco sales, with specific socio-economic regions benefiting the most.
3. Age Restriction (P3): Led to a substantial reduction in tobacco consumption among *young-adult agents*.
4. Combined Approach (P4): Combining all the policies amplified their individual benefits, showcasing the potential of multi-dimensional interventions. Nevertheless, socio-economic disparities persist.

Our policy simulation insights, in alignment with broader literature, underscore that a combined policy approach (like P4) is more effective and promising in reducing tobacco use. However, despite this approach, some socio-economic disparities continue to exist. Our findings strongly advocate for a comprehensive approach to tobacco control.

Chapter 6

Conclusions

In this chapter, we summarise the key contributions of this thesis and explore future research directions.

6.1 Summary of contributions

Tobacco use is one of the leading public health threats, creating a complex, multifaceted landscape influenced by a various factors. Among these factors, interactions between individuals and with the social environment play a pivotal role in determining an individual's smoking behaviour. The novelty of this thesis lies in harnessing the social contagion aspect of smoking through the use of Agent-Based Models (ABMs) to offer new perspectives and potential interventions for curbing tobacco use.

In chapter 2 and 3 we focus on tobacco use as a social contagion on networks, influenced by complex interactions among individuals. To capture this complexity, we introduce an agent-based network model in Chapter 2. This model accounts for multiple smoking processes including initiation, cessation, and relapse through interactions between individuals and spontaneous terms. Our findings underscore the critical role that network degree plays in the spread of smoking behaviours. Furthermore, we find that synthetic networks with comparable estimated degrees can serve as effective substitutes when complete network data is unavailable. Leveraging these insights into the contagious nature of smoking, Chapter 3 focuses on the development of new, network-based intervention strategies for policymakers. These strategies aim to enhance the existing targeting mechanisms employed by smoking cessation clinics. We introduce a simple heuristic grounded in the "friendship paradox"—the notion that your friends, on average, have more friends than you do. We show that this heuristic improves the effectiveness of interventions, leading to a more equitable reduction in smoking prevalence across the network. Importantly, the proposed intervention strategy is not only equitable but also cost-neutral, requiring no changes to existing infrastructure or additional training for clinic staff.

Smoking initiation often occurs at a young age and is usually a gradual process influenced by multiple factors, including parental guidance, peer pressure, susceptibility to addiction, and access to tobacco products. In Chapter 4, we introduce a heterogeneous spatial agent-based model which captures these factors influencing smoking initiation. The model is designed to encapsulate interactions among tobacco users aged 15-24, as well as interactions between these users and tobacco products and the spatial environment. Drawing extensively on data from Scotland, the model serves as a simulation test-bed for evaluating a range of policies across various metrics including prevalence rates, tobacco consumption levels, and socio-economic implications. In Chapter 5 we apply the heterogeneous spatial agent-based model developed in Chapter 4 to evaluate the impact of a variety of tobacco control policies. Specifically, we test the effects of raising tobacco product prices, banning tobacco sales near schools, increasing the minimum legal age for tobacco purchases to 21, and implementing a multi-faceted strategy that combines all these measures. Our analysis indicates that while each policy offers distinct advantages on its own, a comprehensive approach addressing multiple aspects of tobacco use achieves the most substantial reductions in tobacco consumption. Importantly, such an integrated strategy also alleviates the socio-economic disparities often worsened by stand-alone policies. This chapter underscores the versatility of agent-based models as indispensable tools for policymakers, enabling them to simulate a wide range of ‘what-if’ scenarios.

6.2 Discussions

The tobacco use and regulation represents a multifaceted landscape, characterised by diverse interactions at individual, community, and policy levels. In this thesis, we have employed a complex systems approach to navigate this intricate terrain. Specifically, we have developed agent-based models to simulate empirically observed behaviours, aiming to identify the key factors influencing these behaviours. One such model explores interactions between individuals and the phenomenon of social contagion related to smoking behaviour, while another model focuses on the interactions between individuals and the retail environment.

Agent-based models are particularly suitable for investigating ‘what-if’ scenarios, providing policymakers with robust tools for understanding potential outcomes of various interventions. However, it is important to clarify the role of ABMs in the context of tobacco control. As Elsenbroich and Badham (2023) argue, ABMs should not be seen as predictive tools, but rather as means to generate plausible futures based on theoretical assumptions and empirical data. ABMs offer several key advantages for policy analysis in the context of tobacco control:

1. ABMs can represent complex systems with interacting heterogeneous individuals, capturing emerging phenomenon that may be missed with aggregate models. This is particularly relevant for tobacco use, where individual behaviours are influenced by social connections, environmental factors and policy interventions.

2. ABMs combine theoretical understanding through model rules with empirical data via calibration and validation mechanisms, thus providing a balanced approach for exploring ’what-if’ scenarios. In our models, we have integrated social contagion and spatial dynamics with smoking behaviours for understanding policy impacts.

3. The explicit representation of model rules and interactions allow for scrutiny and discussion of assumptions which are important for policy making.

4. Rather than predicting a single outcome, ABMs allow us to explore multiple scenarios, enhancing the understanding of system dynamics and potential policy impacts.

Rather than aiming to predict the future, our models serve to indicate the range of possibilities that the current environment could yield. This rationale underpins our methodology, where we compare model outcomes under different policy conditions and run multiple simulations to gauge a range of potential directions.

Although the idea of an all-encompassing model is theoretically appealing, practical constraints require a more focused approach. Developing an agent-based model necessitates in-depth knowledge of the system being modelled, or at least active engagement with major stakeholders to identify crucial factors. Recognising these constraints, our thesis has selectively targeted its modelling capabilities. We have actively collaborated with Public Health Scotland to understand the tobacco environment and have also sought insights from other researchers in public health. These engagements have not only aided in the practical aspects of policy recommendations but have also emphasised the importance of stakeholder engagement for a comprehensive understanding of the system.

6.3 Future Directions

Methodological advancements An immediate next step would be to integrate the two different models that we have developed. This would involve a multi-network approach that captures both spatial and social network-based interactions. Doing so will allow us to more effectively capture the synergy between spatial and social influences on tobacco use. Additionally, the model could be enriched by incorporating a cessation loop, offering a more nuanced depiction of smoking cessation behaviours.

Expanding to e-cigarettes Given the growing prevalence of e-cigarette use among teenagers (Tehrani et al., 2022; Wang et al., 2020) — a trend that is notably on the rise even as traditional cigarette use declines — it is imperative to expand our model to include this new form of tobacco use. Incorporating e-cigarettes will enable us to investigate its unique dynamics, both in isolation and in conjunction with traditional cigarettes, as well as the policy implications that are becoming increasingly relevant. This expansion is crucial for understanding and addressing the shifting landscape of tobacco use, especially among younger generations who are more susceptible to e-cigarette adoption.

Policy experimentation Our framework could also benefit from the testing of more novel policies. For instance, the idea of annually raising the legal age for tobacco purchase offers a unique avenue for study (Kirby, 2023; Parliament, 2022). Likewise, further research examining the impact of varying retailer densities could yield crucial insights into effective regulation strategies (Ackerman et al., 2017).

Stakeholder engagement for better mechanisms One of the major limitations of the current models is the paucity of high-quality network data. Engaging more closely with policymakers and researchers will not only contribute to model refinement but also emphasise the urgent need for better data collection on social networks affecting tobacco use and similar social contagion. Additionally, stakeholders can provide practical insights into policy implementation challenges, allowing for more realistic modelling of interventions. This multi-faceted engagement will enhance both the accuracy and relevance of our models for real-world policy decisions in tobacco control.

Promoting complex systems thinking among policymakers Lastly, continued engagement with policymakers is essential to advocate for the adoption of a complex systems perspective. This would encourage a multifaceted approach to policy formulation, underlining the importance of understanding the intricacies of the tobacco use landscape for the creation of more effective policies.

Through this thesis, we have utilised agent-based models to explore the regulatory landscape of tobacco use from a complex systems perspective. Looking ahead, we outline several future directions, as detailed earlier in the text. Most crucially, we advocate for the adoption of complex systems thinking by policymakers to craft more effective and holistic strategies in the realm of tobacco regulation. This research serves as a substantive contribution to our collective understanding of tobacco use as a complex system. We intend to build on this work to facilitate more informed and impactful policy decisions.

Appendix A

Appendix

A.1 ODE model

Compartmental models are differential equation models in which the entire population is divided in to different compartments. In our case, each individual in the population is part of one of the three compartments: never-smoker, smoker or quitter. Similar to the popular SIR model, our model also divides the population into three compartments. However, unlike the traditional SIR model, in our case, both infection (smoking initiation in our case) and recovery (quitting in our case) are contagious. In addition we consider relapse and never-smokers influencing smokers too. These are governed by the following equations.

$$\dot{N} = -\frac{\beta_{N,S\to S,S}}{n}NS - \delta_{N\to S}S \tag{A.1}$$

$$\begin{aligned}\dot{S} = {} & \frac{\beta_{N,S\to S,S}}{n}NS - \frac{\beta_{S,N\to Q,N}}{n}SN - \frac{\beta_{S,Q\to Q,Q}}{n}SQ \\ & + \frac{\beta_{Q,S\to S,S}}{n}QS + N\delta_{N\to S} + Q\delta_{Q\to S} - S\delta_{S\to Q}\end{aligned} \tag{A.2}$$

$$\begin{aligned}\dot{Q} = {} & \frac{\beta_{S,N\to Q,N}}{n}SN + \frac{\beta_{S,Q\to Q,Q}}{n}SQ - \frac{\beta_{Q,S\to S,S}}{n}QS \\ & + N\delta_{N\to S} + Q\delta_{Q\to S} - S\delta_{S\to Q}\end{aligned} \tag{A.3}$$

Network	Parameters	Average degree
BA	$m_{BA} = 2$	4
ER	$p_{ER} = 0.003$	3.1
SW	$k = 4$, $p = 0.3$	4
FHS	degree distribution	3.0
LFR	$\tau_1 = 2.5$, $\tau_2 = 1.5$, $\mu = 0.6$, $< k >= 3$	3.0

Table A.1: Model parameters involved in network generation

The equations above models the temporal evolution of each of the compartments: never-smokers $N(t)$, smokers $S(t)$, and quitters $Q(t)$. At any point in time, the total population $n = N + S + Q$. The interaction parameters are given by β's with the subscripts showing the interacting states and the spontaneous terms are given by δ's with the subscripts showing the direction of flow of individuals between compartments.

As mentioned in the main section of the paper, the ODE model assumes a perfect mixing situation where any individual can interact with any other individual in the population.

A.2 Parameters used for network generation

Table A.1 shows the parameter values used to generate networks in each run of the ABM. Each of the parameters were chosen, so that the average degree of the network, was as close as possible to the average degree observed in the FHS network. For LFR networks, we used the standard parameters used for bench-marking (Lancichinetti et al., 2008) along with added condition of the average degree.

A.3 Effect of parameter $\beta_{Q,S\rightarrow S,S}$ in ABM on FC in the UK

Figure A.1 shows how the other parameters change in each of the network, when the parameter $\beta_{Q,S\rightarrow S,S}$ changes. We see that the parameter $\beta_{Q,S\rightarrow S,S}$ (interaction based relapse) doesn't have any effect on the other parameters in the 100 best fit parameters. This along with the fact that the range of parameters (range = 0 to 1) we see in the best

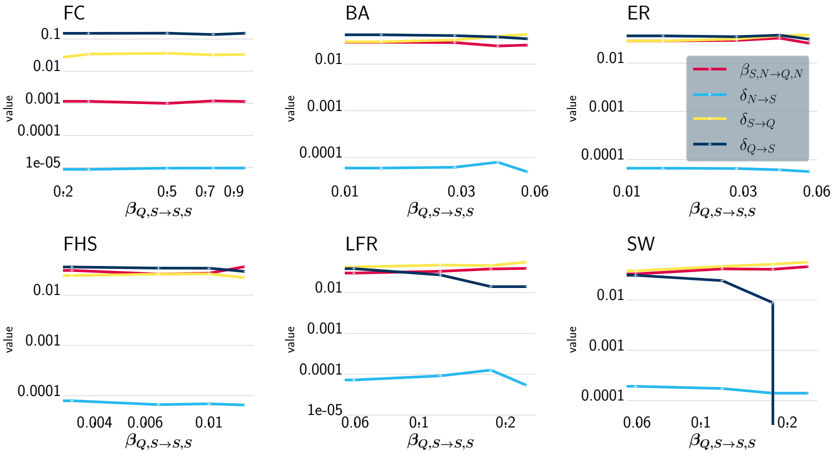


Figure A.1: Effect of parameter $\beta_{Q,S\to S,S}$ on other parameters in the 100 best-fit parameter combinations. In clockwise direction, we have FC, BA, ER, SW, LFR and FHS networks.

100 parameters shows us that changing the value of the $\beta_{Q,S\to S,S}$ doesn't affect the accuracy of the model in mimicking the empirical data for the ABM on FC network in the UK. However, this is not observed in the case of FC network in the US.

In the case of other networks, $\beta_{Q,S\to S,S}$ does not make much of a difference except in the case of SW network ($\delta_{Q\to S}$). However the value of $\beta_{Q,S\to S,S}$ doesn't span the entire range of values available, so we can't conclusively say that parameter $\beta_{Q,S\to S,S}$ is redundant in the other cases.

A.4 Network analysis

Figure A.2 shows the average network characteristics of the 1000 networks used for running the ABM on all network topologies other than the FC network. The first row depicts the degree distribution averaged over 1000 iterations. Not surprisingly, only the BA and LFR networks produced networks with a high degree heterogeneity, while the FHS, ER and SW networks have a similar range of degrees in distribution. The second row shows the distribution of average degree, which by design is around the value $\langle k\rangle = 3$ in ER, FHS and LFR. While, in the case of BA and SW it is around the value of 4.0. The third row exhibits the number of edges distribution, with comparable ranges for FHS, ER and LFR agreeing with their similar average degrees. The fourth

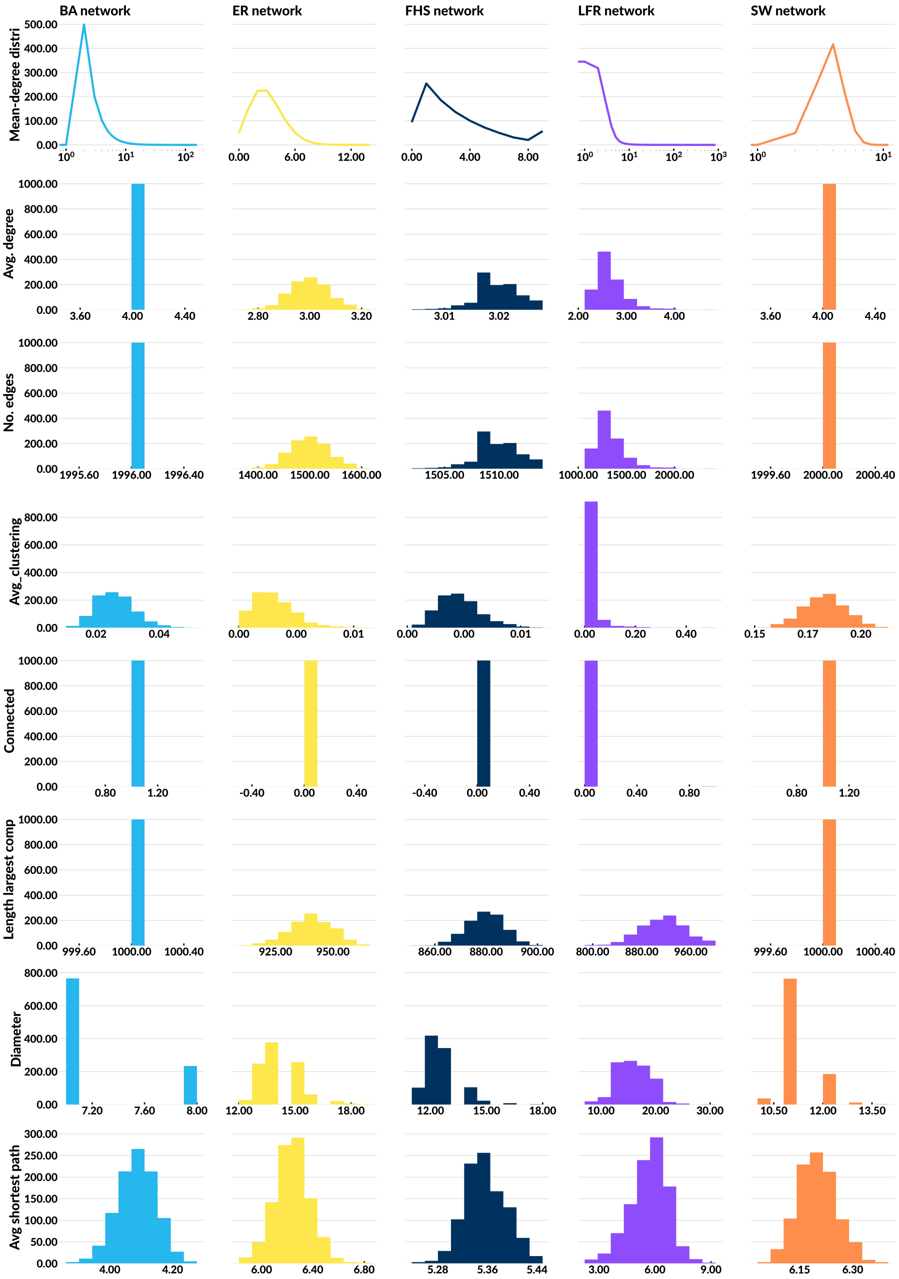


Figure A.2: The figure shows the average network characteristics of the 1000 simulated networks used for validation for each network topology. Each column represents a different network topology, while each row gives the corresponding network characteristic.

row presents the average clustering coefficient, quantifying node clustering tendencies. The fifth row indicates network connectedness - BA and SW are fully connected, unlike the others. This can be seen in the sixth row as well, which shows the largest connected component size, with BA and SW comprising a single giant component versus disconnected subnetworks for the rest. The seventh and eighth rows display the diameter and average shortest path length respectively. Notably, BA has the shortest diameter and path length, while ER and FHS have the smallest clustering. Overall, these statistics suggest that ER possesses a greater degree of similarity to the real-world FHS network than the other networks.

A.5 Sensitivity analysis

Figure A.3 describes the effect of individual parameters on the temporal dynamics of the ABM on FC network around the best fit parameters. From Figure A.3 we can see that by increasing the parameters $\beta_{Q,S\to S,S}$, $\delta_{N\to S}$, and $\delta_{Q\to S}$, the S population increases while the Q population decreases. These are the parameters that directly increase S population size, and thus the result can be logically reasoned from the model structure. At the same time, by increasing the parameters $\delta_{S\to Q}$ and $\beta_{S,N\to Q,N}$, the value of Q curve, and the value of S decreases. Only $\delta_{N\to S}$ affects the N curve, which is the parameter directly involved with smoking initiation. Since the N population is not affected much, the S and Q curves are largely symmetrical in all other cases.

Just as in Figure A.3, the other networks also behave similarly in terms of how each parameter affects the population dynamics. This can be seen in Figures A.4, A.5, A.6, A.7 and A.8. However, the degree to which each parameter affects the dynamics differs, especially between FC and the other networks. Our results from the Sensitivity analysis show that the FC network is more sensitive to lower values of parameters for all parameters except $\delta_{Q\to S}$. At higher values of most of these parameters, the population behaviour saturates and converges to a similar value.

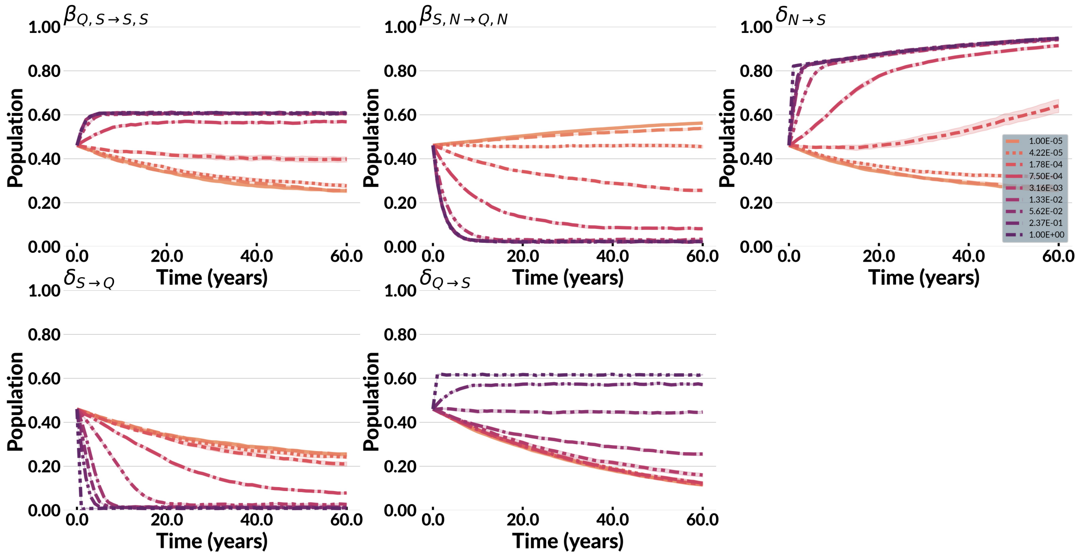


Figure A.3: Sensitivity analysis of parameters around the best-fit parameters for the ABM on FC. From left to right, row by row, the effect of increasing parameters $\beta_{Q,S\to S,S}$, $\beta_{N,S\to S,S}$, $\delta_{N\to S}$, $\delta_{S\to Q}$, and $\delta_{Q\to S}$ on the smoker population dynamics is displayed. Each parameter is increased from 0 to 1 (10 values) logarithmically, while all other parameters are kept at the minimum MSE parameter combination.

In the case of all the other networks, the behaviour displayed on changing parameters around the best-fit combination is very similar. This is probably due to each of these networks having similar average degrees.

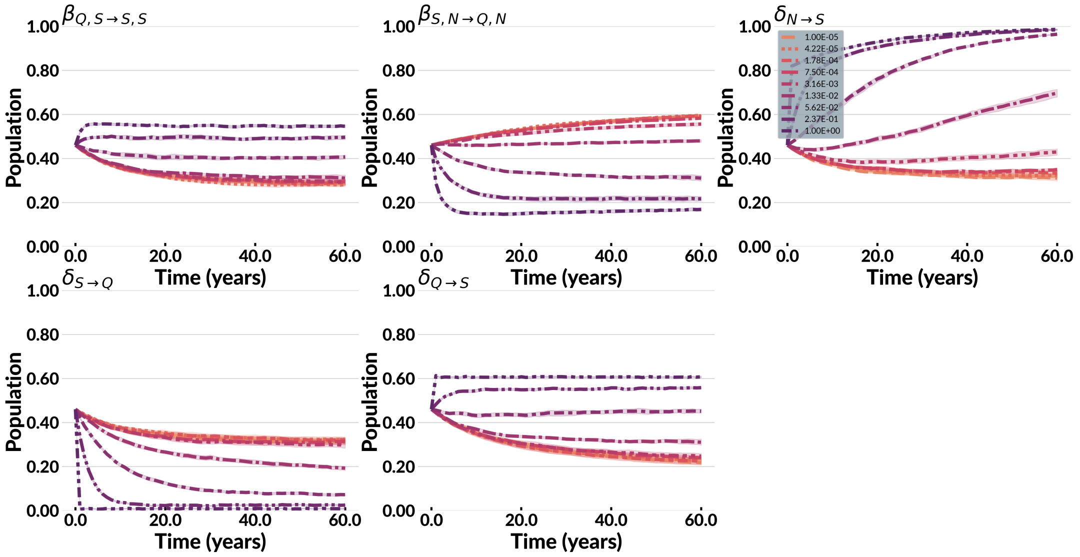


Figure A.4: Sensitivity analysis of parameters around the best fit parameters for the ABM on BA. From left to right, row by row, the effect of increasing parameters $\beta_{Q,S\to S,S}$, $\beta_{N,S\to S,S}$, $\delta_{N\to S}$, $\delta_{S\to Q}$, and $\delta_{Q\to S}$ on the smoker population dynamics is displayed. Each parameter is increased from 0 to 1 (10 values) logarithmically, while all other parameters are kept at the minimum MSE parameter combination.

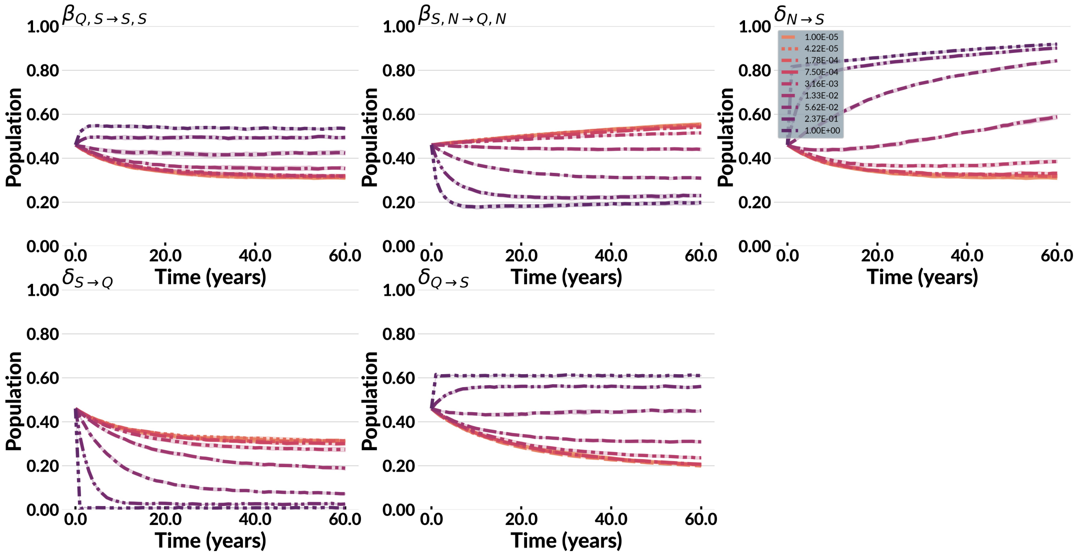


Figure A.5: Sensitivity analysis of parameters around the best fit parameters for the ABM on ER. From left to right, row by row, the effect of increasing parameters $\beta_{Q,S\to S,S}$, $\beta_{N,S\to S,S}$, $\delta_{N\to S}$, $\delta_{S\to Q}$, and $\delta_{Q\to S}$ on the smoker population dynamics is displayed. Each parameter is increased from 0 to 1 (10 values) logarithmically, while all other parameters are kept at the minimum MSE parameter combination.

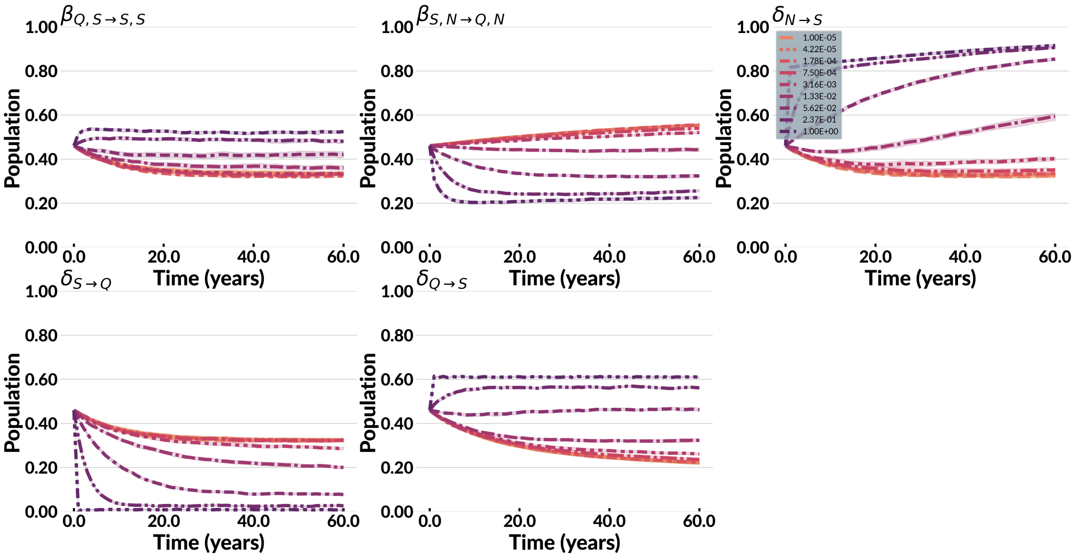


Figure A.6: Sensitivity analysis of parameters around the best fit parameters for the ABM on FHS. From left to right, row by row, the effect of increasing parameters $\beta_{Q,S\to S,S}$, $\beta_{N,S\to S,S}$, $\delta_{N\to S}$, $\delta_{S\to Q}$, and $\delta_{Q\to S}$ on the smoker population dynamics is displayed. Each parameter is increased from 0 to 1 (10 values) logarithmically, while all other parameters are kept at the minimum MSE parameter combination.

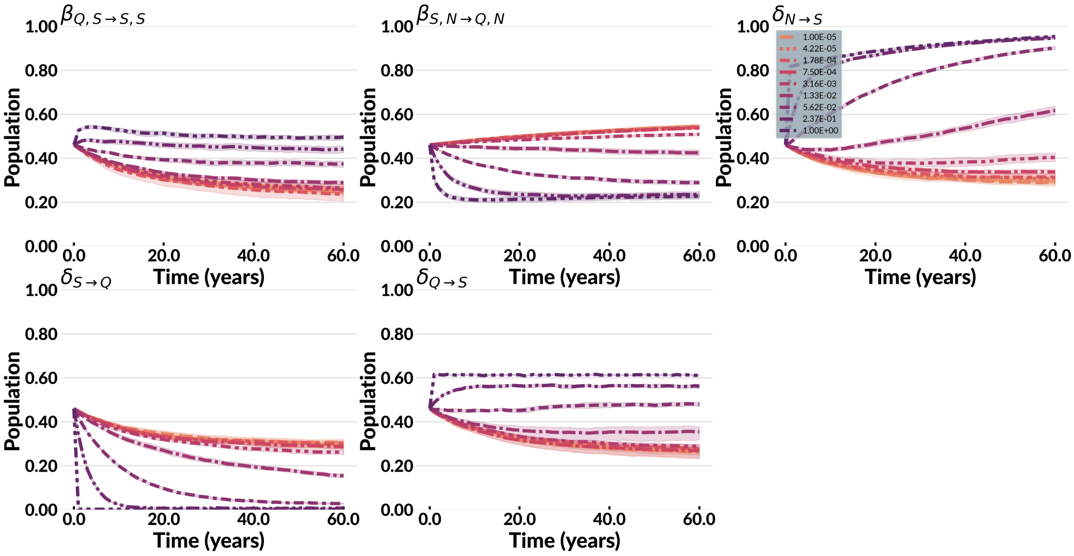


Figure A.7: Sensitivity analysis of parameters around the best fit parameters for the ABM on LFR. From left to right, row by row, the effect of increasing parameters $\beta_{Q,S\to S,S}$, $\beta_{N,S\to S,S}$, $\delta_{N\to S}$, $\delta_{S\to Q}$, and $\delta_{Q\to S}$ on the smoker population dynamics is displayed. Each parameter is increased from 0 to 1 (10 values) logarithmically, while all other parameters are kept at the minimum MSE parameter combination.

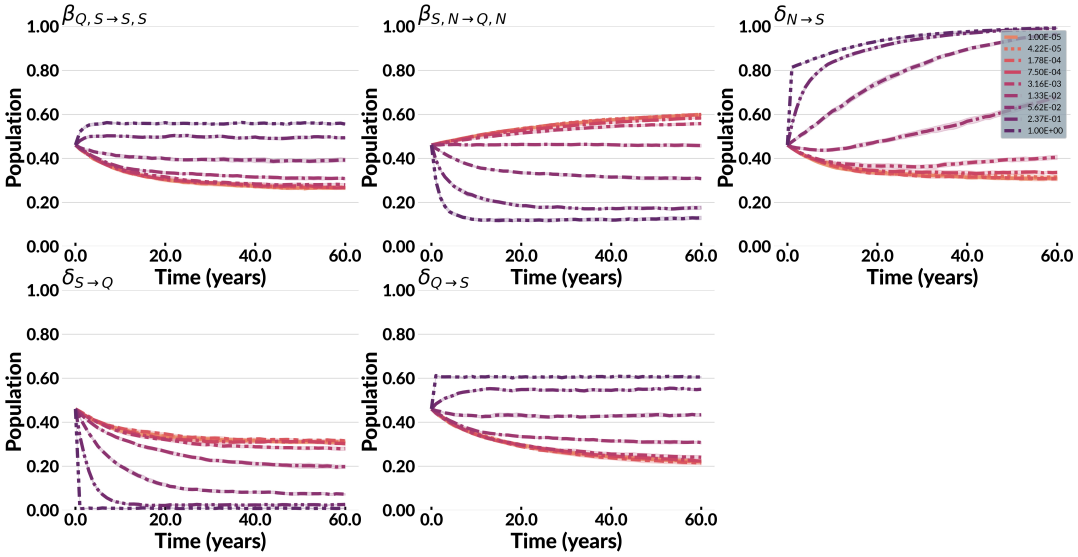


Figure A.8: Sensitivity analysis of parameters around the best fit parameters for the ABM on SW. From left to right, row by row, the effect of increasing parameters $\beta_{Q,S\to S,S}$, $\beta_{N,S\to S,S}$, $\delta_{N\to S}$, $\delta_{S\to Q}$, and $\delta_{Q\to S}$ on the smoker population dynamics is displayed. Each parameter is increased from 0 to 1 (10 values) logarithmically, while all other parameters are kept at the minimum MSE parameter combination.

A.6 Effect of number of agents

Figure A.9 describes the population dynamics of the ABM on LFR network when the total population size is varied (and the best fit parameters). The results demonstrate consistent S and Q prevalence dynamics across different populations sizes. As expected, variance increases at $n = 100$ due to greater stochastic effects with fewer agents. However, the overall behaviour aligns well for all population sizes, confirming model robustness to population size.

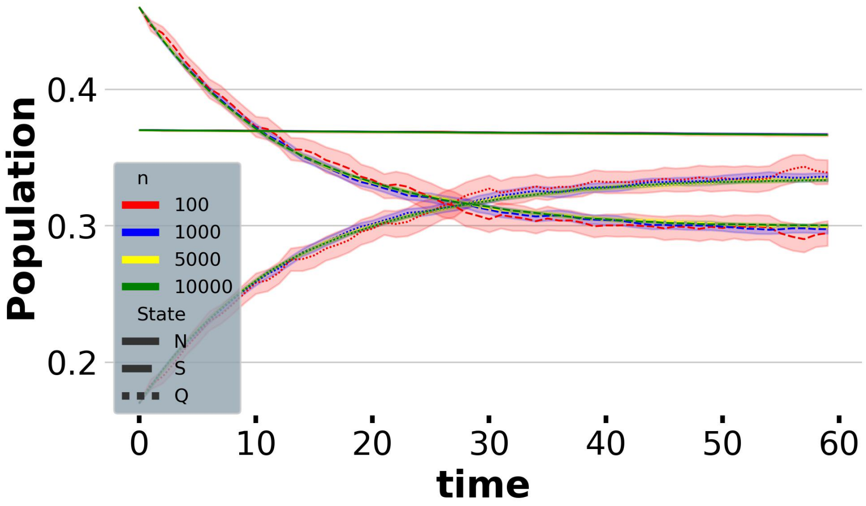


Figure A.9: Sensitivity analysis on population size for the ABM on the LFR network. The total agent population n was varied as 100, 1000, 5000 and $10,000$ to assess effects on model dynamics. The mean simulated smoker (S) and quitter (Q) prevalence over time is shown for each n, with shaded regions indicating 95% confidence intervals across 1000 runs. From the figure we see that the model dynamics does not change based on the population size.

Bibliography

Arnab Acharya, Kathryn Angus, Samira Asma, Douglas W Bettcher, Kenneth Blackman, Evan Blecher, Ron Borland, Christina Ciecierski, Meishan Cui, Vera Luiza da Costa e Silva, et al. The economics of tobacco and tobacco control. Technical report, 2016.

Amy Ackerman, Alexis Etow, Sara Bartel, and Kurt M Ribisl. Reducing the density and number of tobacco retailers: policy solutions and legal issues. *Nicotine & Tobacco Research*, 19(2):133–140, 2017.

Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *nature*, 401(6749):130–131, 1999.

Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini Jr, Stefano Merler, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. *Nature Human Behaviour*, 4(9):964–971, 2020.

Alberto Aleta, David Martín-Corral, Michiel A Bakker, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini Jr, et al. Quantifying the importance and location of sars-cov-2 transmission events in large metropolitan areas. *Proceedings of the National Academy of Sciences*, 119(26):e2112182119, 2022.

Fatma Romeh M Ali, Israel T Agaku, Saida R Sharapova, Elizabeth A Reimels, and David M Homa. Peer reviewed: onset of regular smoking before age 21 and subsequent nicotine dependence and cessation behavior among us adult smokers. *Preventing chronic disease*, 17, 2020.

Roy M Anderson, B Anderson, and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

Philip Emeka Anyanwu, Peter Craig, Srinivasa Vittal Katikireddi, and Michael James Green. Impact of uk tobacco control policies on inequalities in youth smoking uptake: a natural experiment study. *Nicotine and Tobacco Research*, 22(11):1973–1980, 2020.

W Brian Arthur. Complexity and the economy. *science*, 284(5411):107–109, 1999.

ASH Scotland. Smoking in scotland where are we now. *Edinburgh, Scotland: ASH Scotland*, 2014.

Hutan Ashrafian, Tania Toma, Leanne Harling, Karen Kerr, Thanos Athanasiou, and Ara Darzi. Social networking strategies that aim to reduce obesity have achieved significant although modest results. *Health affairs*, 33(9):1641–1647, 2014.

Jennifer Badham, Edmund Chattoe-Brown, Nigel Gilbert, Zaid Chalabi, Frank Kee, and Ruth F Hunter. Developing agent-based models of complex health behaviour. *Health & place*, 54:170–177, 2018.

Jennifer Badham, Frank Kee, and Ruth F Hunter. Network structure influence on simulated network interventions for behaviour change. *Social Networks*, 64:55–62, 2021.

Emily Banks, Grace Joshy, Marianne F. Weber, Bette Liu, Robert Grenfell, Sam Egger, Ellie Paige, Alan D. Lopez, Freddy Sitas, and Valerie Beral. Tobacco smoking and all-cause mortality in a large Australian cohort study: Findings from a mature epidemic with current low smoking prevalence. *BMC Medicine*, 13(1): 38, feb 2015. ISSN 17417015. doi: 10.1186/s12916-015-0281-z. URL http://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-015-0281-z.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

Linda Bauld, Ken Judge, and Stephen Platt. Assessing the impact of smoking cessation services on reducing health inequalities in england: observational study. *Tobacco control*, 16(6):400–404, 2007.

Linda Bauld, Kirsten Bell, Lucy McCullough, Lindsay Richardson, and Lorraine Greaves. The effectiveness of nhs smoking cessation services: a systematic review. *Journal of Public Health*, 32(1):71–82, 2010.

Kunal Bhattacharya, Asim Ghosh, Daniel Monsivais, Robin IM Dunbar, and Kimmo Kaski. Sex differences in social focus across the life cycle in humans. *Royal Society open science*, 3(4):160097, 2016.

Tony Blakely and Coral Gartner. Tobacco taxes have mixed effects on socioeconomic disparities. *The Lancet Public Health*, 4(12):e595–e596, 2019.

Per Block, Marion Hoffman, Isabel J Raabe, Jennifer Beam Dowd, Charles Rahal, Ridhi Kashyap, and Melinda C Mills. Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world. *Nature Human Behaviour*, 4 (6):588–596, 2020.

David J Blok, Sake J de Vlas, Pepijn van Empelen, and Frank J van Lenthe. The role of smoking in social networks on smoking cessation and relapse among adults: A longitudinal study. *Preventive medicine*, 99:105–110, 2017.

Richard J Bonnie, Kathleen Stratton, Leslie Y Kwan, et al. *Public health implications of raising the minimum age of legal access to tobacco products*. National Academies Press Washington, DC, 2015.

Naomi Breslau and Edward L Peterson. Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. *American journal of public health*, 86(2):214–220, 1996.

Naomi Breslau, Nancy Fenn, and Edward L Peterson. Early smoking initiation and nicotine dependence in a cohort of young adults. *Drug and alcohol dependence*, 33 (2):129–137, 1993.

Leonie S Brose, Robert West, Máirtín S McDermott, Jennifer A Fidler, Emma Croghan, and Andy McEwen. What makes for an effective stop-smoking service? *Thorax*, 66 (10):924–926, 2011.

Jamie Brown, Daniel Kotz, Susan Michie, John Stapleton, Matthew Walmsley, and Robert West. How effective and cost-effective was the national mass media smoking cessation campaign 'stoptober'? *Drug and alcohol dependence*, 135:52–58, 2014a.

Tamara Brown, Stephen Platt, and Amanda Amos. Equity impact of interventions and policies to reduce smoking in youth: systematic review. *Tobacco control*, 23(e2): e98–e105, 2014b.

Albert J Burgess-Hull, Linda J Roberts, Megan E Piper, and Timothy B Baker. The social networks of smokers attempting to quit: An empirically derived and validated classification. *Psychology of Addictive Behaviors*, 32(1):64, 2018.

Jennifer Cantrell, Jennifer L Pearson, Andrew Anesetti-Rothermel, Haijun Xiao, Thomas R Kirchner, and Donna Vallone. Tobacco retail outlet density and young adult tobacco initiation. *Nicotine & Tobacco Research*, 18(2):130–137, 2015.

George T Cantwell, Alec Kirkley, and MEJ Newman. The friendship paradox in real and model networks. *Journal of Complex Networks*, 9(2):cnab011, 2021.

Fiona Caryl, Niamh K Shortt, Jamie Pearce, Garth Reid, and Richard Mitchell. Socioeconomic inequalities in children's exposure to tobacco retailing based on individual-level gps data in scotland. *Tobacco Control*, 29(4):367–373, 2020.

Fiona M Caryl, Jamie Pearce, Garth Reid, Richard Mitchell, and Niamh K Shortt. Simulating the density reduction and equity impact of potential tobacco retail control policies. *Tobacco control*, 30(e2):e138–e143, 2021.

Carlos Castillo-Garsow, Guarionex Jordan-Salivia, Ariel Rodriguez-Herrera, et al. Mathematical models for the dynamics of tobacco use, recovery and relapse. *Technical Report Series*, 1997.

CDC. Youth risk behavior surveillance system (yrbss). *Centers for Disease Control and Prevention (CDC)*, 2016.

Magdalena Cerdá, Melissa Tracy, Katherine M Keyes, and Sandro Galea. To treat or to prevent?: Reducing the population burden of violence-related post-traumatic stress disorder. *Epidemiology (Cambridge, Mass.)*, 26(5):681, 2015.

Dingding Chao, Hideki Hashimoto, and Naoki Kondo. Dynamic impact of social stratification and social influence on smoking prevalence by gender: an agent-based model. *Social Science & Medicine*, 147:280–287, 2015.

Chartered Trading Standards Institute 2015. Enhancing local tobacco control: preventing under age sales of tobacco, 2015.

Laurie Chassin, Clark C Presson, Steven J Sherman, and Debra A Edwards. The natural history of cigarette smoking: predicting young-adult smoking outcomes from adolescent smoking patterns. *Health psychology*, 9(6):701, 1990.

Seulggie Choi, Jooyoung Chang, Kyuwoong Kim, Sang Min Park, and Kiheon Lee. Effect of smoking cessation and reduction on the risk of cancer in korean men: a population based study. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, 50(4):1114–1120, 2018.

Won S Choi, Kari Jo Harris, Kolawole Okuyemi, and Jasjit S Ahluwalia. Predictors of smoking initiation among college-bound high school students. *Annals of Behavioral Medicine*, 26(1):69–74, 2003.

Nicholas A Christakis and James H Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

Francesca Colaiori and Claudio Castellano. Interplay between media and social influence in the collective behavior of opinion dynamics. *Physical Review E*, 92(4): 042815, 2015.

CRESH Research Group. Neighbourhood alcohol and tobacco environments in scotland, 2023.

Yang Cui, Evelyn L Forget, Yunfa Zhu, Mahmoud Torabi, and Umut Oguzoglu. The effects of cigarette price and the amount of pocket money on youth smoking initiation and intensity in canada. *Canadian Journal of Public Health*, 110:93–102, 2019.

Magdalena Czubek and Surjinder Johal. *Econometric analysis of cigarette consumption in the UK.* Citeseer, 2010.

Thomas Royle Dawber. The framingham study. In *The Framingham Study*. Harvard university press, 2013.

Department of Health. The children and young persons (sale of tobacco etc.) order 2007. *Smoking: prohibition and control*, 2007.

Department of Health and Social Care. Children whose parents smoke are 4 times as likely to take up smoking themselves, 2021.

Department of Transport, United Kingdom. Travel time measures for local 'a' roads: January to december 2021 report, 2022.

Fiona Dobbie, Rosemary Hiscock, Jo Leonardi-Bee, Susan Murray, Lion Shahab, Paul Aveyard, Tim Coleman, Andy McEwen, Hayden McRobbie, Richard Purves, et al. Evaluating long-term outcomes of nhs stop smoking services (elons): a prospective cohort study. *Health Technology Assessment (Winchester, England)*, 19(95):1, 2015.

Lori Dorfman and Lawrence Wallack. Advertising health: the case for counter-ads. *Public health reports*, 108(6):716, 1993.

J Drope, N Schluger, Z Cahn, J Drope, S Hamill, F Islami, A Liber, N Nargis, and M Stoklosa. *The Tobacco Atlas, Sixth edition Jeffrey Drope and Neil W. Schluger, editors*. 2018. ISBN 9781604432572.

Meredith S Duncan, Matthew S Freiberg, Robert A Greevy, Suman Kundu, Ramachandran S Vasan, and Hilary A Tindle. Association of smoking cessation with subsequent risk of cardiovascular disease. *Jama*, 322(7):642–650, 2019.

Corinna Elsenbroich and Jennifer Badham. Negotiating a future that is not like the past. *International Journal of Social Research Methodology*, 26(2):207–213, 2023.

Susan T Ennett, Robert Faris, John Hipp, Vangie A Foshee, Karl E Bauman, Andrea Hussong, and Li Cai. Peer smoking, other peer attributes, and adolescent cigarette smoking: A social network analysis. *Prevention Science*, 9(2):88–98, 2008.

Paul Erdős, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

Shari P Feirman, Allison M Glasser, Shyanika Rose, Ray Niaura, David B Abrams, Lyubov Teplitskaya, and Andrea C Villanti. Computational models used to assess us tobacco control policies. *Nicotine & Tobacco Research*, 19(11):1257–1267, 2017.

Scott L Feld. Why your friends have more friends than you do. *American journal of sociology*, 96(6):1464–1477, 1991.

Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. 2020.

Thomas Ferkol and Dean Schraufnagel. The global burden of respiratory disease. *Annals of the American Thoracic Society*, 11(3):404–406, 2014.

Jennifer A Fidler and Robert West. Changes in smoking prevalence in 16–17-year-old versus older adults following a rise in legal age of sale: findings from an english population study. *Addiction*, 105(11):1984–1988, 2010.

Filippos T Filippidis, Israel T Agaku, and Constantine I Vardavas. The association between peer, parental influence and tobacco product features and earlier age of onset of regular smoking among adults in 27 european countries. *The European Journal of Public Health*, 25(5):814–818, 2015.

Luisa S Flor, Marissa B Reitsma, Vinay Gupta, Marie Ng, and Emmanuela Gakidou. The effects of tobacco control policies on global smoking prevalence. *Nature Medicine*, 27(2):239–243, 2021.

D Gamermann, J Triana-Dopico, and R Jaime. A comprehensive statistical study of metabolic and protein–protein interaction network properties. *Physica A: Statistical Mechanics and its Applications*, 534:122204, 2019.

GBD 2019 Tobacco Collaborators and others. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the global burden of disease study 2019. *Lancet (London, England)*, 397(10292):2337, 2021.

Stephen E Gilman, Richard Rende, Julie Boergers, David B Abrams, Stephen L Buka, Melissa A Clark, Suzanne M Colby, Brian Hitsman, Alessandra N Kazura, Lewis P Lipsitt, et al. Parental smoking and adolescent smoking initiation: an intergenerational perspective on tobacco control. *Pediatrics*, 123(2):e274–e281, 2009.

Anna B Gilmore, Behrooz Tavakoly, Rosemary Hiscock, and Gordon Taylor. Smoking patterns in great britain: the rise of cheap cigarette brands and roll your own (ryo) tobacco. *Journal of public health*, 37(1):78–88, 2015.

Elizabeth A Gilpin and John P Pierce. Trends in adolescent smoking initiation in the united states: is tobacco marketing an influence? *Tobacco Control*, 6(2):122–127, 1997.

Allison M Glasser and Megan E Roberts. Retailer density reduction approaches to tobacco control: a review. *Health & place*, 67:102342, 2021.

Myong-Hyun Go, Harold D Green Jr, David P Kennedy, Michael Pollard, and Joan S Tucker. Peer influence and selection effects on adolescent smoking. *Drug and Alcohol Dependence*, 109(1-3):239–242, 2010.

gohenry. The gohenry youth economy report,2022, 2022.

Martin Gonzalez-Rozada and Giselle Montamat. How raising tobacco prices affects the decision to start and quit smoking: evidence from argentina. *International journal of environmental research and public health*, 16(19):3622, 2019.

Mark Goodchild, Nigar Nargis, and Edouard Tursan d'Espaignet. Global economic cost of smoking-attributable diseases. *Tobacco control*, 27(1):58–64, 2018.

Dennis M Gorman, Jadranka Mezic, Igor Mezic, and Paul J Gruenewald. Agent-based modeling of drinking behavior: a preliminary model and potential applications to theory and practice. *American journal of public health*, 96(11):2055–2060, 2006.

Ashleigh Guillaumier, Billie Bonevski, and Christine Paul. 'cigarettes are priority': a qualitative study of how australian socioeconomically disadvantaged smokers respond to rising cigarette prices. *Health Education Research*, 30(4):599–608, 2015.

Zeena Harakeh and Wilma AM Vollebergh. The impact of active and passive peer influence on young adult smoking: An experimental study. *Drug and alcohol dependence*, 121(3):220–223, 2012.

Sally Haw, Dorothy Currie, Douglas Eadie, Jamie Pearce, Andy MacGregor, Martine Stead, Amanda Amos, Catherine Best, Michael Wilson, Mark Cherrie, et al. The impact of the point-of-sale tobacco display ban on young people in scotland: before-and-after study. *Public Health Research*, 8(1):1–118, 2020.

Lisa Henriksen, Ellen C Feighery, Nina C Schleicher, David W Cowling, Randolph S Kline, and Stephen P Fortmann. Is adolescent smoking related to the density and proximity of tobacco outlets and retail cigarette advertising near schools? *Preventive medicine*, 47(2):210–214, 2008.

Lisa Henriksen, Elizabeth Andersen-Rodgers, Xueying Zhang, April Roeseler, Dennis L Sun, Trent O Johnson, and Nina C Schleicher. Neighborhood variation in the price of cheap tobacco products in california: results from healthy stores for a healthy community. *Nicotine & Tobacco Research*, 19(11):1330–1337, 2017.

Alison L Hill, David G Rand, Martin A Nowak, and Nicholas A Christakis. Emotions as infectious diseases in a large social network: the sisa model. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701):3827–3835, 2010a.

Alison L Hill, David G Rand, Martin A Nowak, and Nicholas A Christakis. Infectious disease modeling of social contagion in networks. *PLOS computational biology*, 6 (11), 2010b.

Rosemary Hiscock and Linda Bauld. Stop smoking services and health inequalities. *National Centre for Smoking Cessation and Training*, 2013.

Rosemary Hiscock, Linda Bauld, Amanda Amos, Jennifer A Fidler, and Marcus Munafò. Socioeconomic status and smoking: a review. *Annals of the New York Academy of Sciences*, 1248(1):107–123, 2012.

Sara C Hitchman, Geoffrey T Fong, Mark P Zanna, James F Thrasher, Janet Chung-Hall, and Mohammad Siahpush. Socioeconomic status and smokers' number of smoking friends: findings from the international tobacco control (itc) four country survey. *Drug and alcohol dependence*, 143:158–166, 2014.

HM Revenue & Customs . Customs (2018),"measuring tax gaps 2018 edition: tax gap estimates for 2016-17", 2018.

HMRC, The govenrment of UK. Changes to tobacco duty rates from 15 march 2023, 2023.

Nathan O Hodas and Kristina Lerman. The simple rules of social contagion. *Scientific reports*, 4(1):1–7, 2014.

Janet Hoek and Katie Smith. A qualitative analysis of low income smokers' responses to tobacco excise tax increases. *International Journal of Drug Policy*, 37:82–89, 2016.

Irving F Hoffman, Carl A Latkin, Polina V Kukhareva, Sergey V Malov, Julia V Batluk, Alla V Shaboltas, Roman V Skochilov, Nicolay V Sokolov, Sergei V Verevochkin, Michael G Hudgens, et al. A peer-educator network hiv prevention intervention among injection drug users: results of a randomized controlled trial in st. petersburg, russia. *AIDS and Behavior*, 17:2510–2520, 2013.

Theodore R Holford, Rafael Meza, Kenneth E Warner, Clare Meernik, Jihyoun Jeon, Suresh H Moolgavkar, and David T Levy. Tobacco control and the reduction in smoking-related premature deaths in the united states, 1964-2012. *Jama*, 311(2): 164–171, 2014.

Elizabeth Hunter, Brian Mac Namee, and John Kelleher. An open-data-driven agent-based model to simulate infectious disease outbreaks. *PloS one*, 13(12):e0208775, 2018.

Ruth F Hunter, Kayla de la Haye, Jennifer M Murray, Jennifer Badham, Thomas W Valente, Mike Clarke, and Frank Kee. Social network interventions for health behaviours and outcomes: A systematic review and meta-analysis. *PLoS medicine*, 16(9):e1002890, 2019.

Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nature communications*, 10(1):2485, 2019.

Ipsos MORI Scotland. Scottish schools adolescent lifestyle and substance use survey, 2020, 2020.

Scottish Government Ipsos MORI Scotland. Scottish household survey, 2016, 2016a.

Scottish Government Ipsos MORI Scotland. Transport and travel in scotland 2016, 2016b.

Martin J Jarvis and Jane Wardle. Social patterning of individual health behaviours: the case of cigarette smoking. *Social determinants of health*, 2:224–37, 1999.

Prabhat Jha and Richard Peto. Global effects of smoking, of quitting, and of taxing tobacco. *New England Journal of Medicine*, 370(1):60–68, 2014.

Prabhat Jha, Richard Peto, Witold Zatonski, Jillian Boreham, Martin J Jarvis, and Alan D Lopez. Social inequalities in male mortality, and in male mortality from smoking: indirect estimation from national death rates in england and wales, poland, and north america. *The Lancet*, 368(9533):367–370, 2006.

Jackie Kazil, David Masad, and Andrew Crooks. Utilizing python for agent-based modeling: The mesa framework. In Robert Thomson, Halil Bisgin, Christopher Dancy, Ayaz Hyder, and Muhammad Hussain, editors, *Social, Cultural, and Behavioral Modeling*, pages 308–317, Cham, 2020. Springer International Publishing. ISBN 978-3-030-61255-9.

Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.

William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 138(834):55–83, 1932.

William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 141(843):94–122, 1933.

Cliff C Kerr, Robyn M Stuart, Dina Mistry, Romesh G Abeysuriya, Katherine Rosenfeld, Gregory R Hart, Rafael C Núñez, Jamie A Cohen, Prashanth Selvaraj, Brittany Hagedorn, et al. Covasim: an agent-based model of covid-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):e1009149, 2021.

Tony Kirby. Uk announces plan to increase legal age of buying tobacco by 1 year every year. *The Lancet Oncology*, 2023.

Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases*, 20(5):553–558, 2020.

James Ladyman, James Lambert, and Karoline Wiesner. What is a complex system? *European Journal for Philosophy of Science*, 3:33–67, 2013.

Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.

Carl Latkin, Deborah Donnell, Ting-Yuan Liu, Melissa Davey-Rothwell, David Celentano, and David Metzger. The dynamic relationship between social norms and behaviors: the results of an hiv prevention network intervention for injection drug users. *Addiction*, 108(5):934–943, 2013.

Carl A Latkin and Amy R Knowlton. Social network assessments and interventions for health behavior change: a critical review. *Behavioral Medicine*, 41(3):90–97, 2015.

S Lee, E Jung, and Carlos Castillo-Chavez. Optimal control intervention strategies in low-and high-risk problem drinking populations. *Socio-Economic Planning Sciences*, 44(4):258–265, 2010.

Jo Leonardi-Bee, Mirriam Lisa Jere, and John Britton. Exposure to parental and sibling smoking and the risk of smoking uptake in childhood and adolescence: a systematic review and meta-analysis. *Thorax*, 66(10):847–855, 2011.

Douglas E Levy, Susan Regan, Giselle K Perez, Alona Muzikansky, Emily R Friedman, Julia Rabin, Nancy A Rigotti, Jamie S Ostroff, and Elyse R Park. Cost-effectiveness of implementing smoking cessation interventions for patients with cancer. *JAMA Network Open*, 5(6):e2216362–e2216362, 2022.

James Lightwood and Stanton A Glantz. Smoking behavior and healthcare expenditure in the united states, 1992–2009: Panel data estimates. *PLoS medicine*, 13(5): e1002020, 2016.

Dean R Lillard, Eamon Molloy, and Andrew Sfekas. Smoking initiation and the iron law of demand. *Journal of health economics*, 32(1):114–127, 2013.

Jiaying Liu, Siman Zhao, Xi Chen, Emily Falk, and Dolores Albarracín. The influence of peer behavior as a function of social and cultural closeness: A meta-analysis of normative influence on adolescent smoking initiation and continuation. *Psychological bulletin*, 143(10):1082, 2017.

Douglas Luke, Sarah Moreland-Russell, Amy Sorg, Heidi Walsh, et al. Point-of-sale report to the nation: the tobacco retail and policy landscape. 2014.

Douglas A Luke, Ross A Hammond, Todd Combs, Amy Sorg, Matt Kasman, Austen Mack-Crane, Kurt M Ribisl, and Lisa Henriksen. Tobacco town: computational modeling of policy options to reduce tobacco retailer density. *American journal of public health*, 107(5):740–746, 2017.

David M Lydon, Stephen J Wilson, Amanda Child, and Charles F Geier. Adolescent brain maturation and smoking: what we know and where we're headed. *Neuroscience & Biobehavioral Reviews*, 45:323–342, 2014.

Louise Marsh, Pavla Vaneckova, Lindsay Robertson, Trent O Johnson, Crile Doscher, Ilana G Raskind, Nina C Schleicher, and Lisa Henriksen. Association between density and proximity of tobacco retail outlets with smoking: a systematic review of youth studies. *Health & Place*, 67:102275, 2021.

Russell Seth Martins, Muhammad Umer Junaid, Muhammad Sharjeel Khan, Namrah Aziz, Zoha Zahid Fazal, Mariam Umoodi, Fatima Shah, and Javaid Ahmed Khan. Factors motivating smoking cessation: a cross-sectional study in a lower-middle-income country. *BMC Public Health*, 21(1):1–11, 2021.

Darren Mays, Stephen E Gilman, Richard Rende, George Luta, Kenneth P Tercyak, and Raymond S Niaura. Parental smoking exposure and adolescent smoking trajectories. *Pediatrics*, 133(6):983–991, 2014.

Liesbeth Mercken, Tom AB Snijders, Christian Steglich, and Hein de Vries. Dynamics of adolescent friendship networks and smoking behavior: Social network analyses in six european countries. *Social science & medicine*, 69(10):1506–1514, 2009.

Christopher Millett, John Tayu Lee, Daniel C Gibbons, and Stanton A Glantz. Increasing the age for the legal purchase of tobacco in england: impacts on socio-economic disparities in youth smoking. *Thorax*, 66(10):862–865, 2011.

Ritesh Mistry, Mangesh Pednekar, Sharmila Pimple, Prakash C Gupta, William J McCarthy, Lalit J Raute, Minal Patel, and Surendra S Shastri. Banning tobacco sales and advertisements near educational institutions may reduce students' tobacco use risk: evidence from mumbai, india. *Tobacco control*, 24(e1):e100–e107, 2015.

Ute Mons, Aysel Müezzinler, Carolin Gellert, Ben Schöttker, Christian C Abnet, Martin Bobak, Lisette de Groot, Neal D Freedman, Eugène Jansen, Frank Kee, et al. Impact of smoking and smoking cessation on cardiovascular events and mortality among older adults: meta-analysis of individual participant data from prospective cohort studies of the chances consortium. *bmj*, 350, 2015.

Cristopher Moore and Mark EJ Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678, 2000.

Shane T Mueller and Yin-Yin Sarah Tan. Cognitive perspectives on opinion dynamics: The role of knowledge in consensus formation, opinion divergence, and group polarization. *Journal of Computational Social Science*, 1:15–48, 2018.

Rachael L Murray, Linda Bauld, Lucy E Hackshaw, and Ann McNeill. Improving access to smoking cessation services for disadvantaged groups: a systematic review. *Journal of Public Health*, 31(2):258–277, 2009.

Allison E Myers, Marissa G Hall, Lisa F Isgett, and Kurt M Ribisl. A comparison of three policy approaches for tobacco retailer reduction. *Preventive medicine*, 74: 67–73, 2015.

Gera E Nagelhout, Dianne de Korte-de Boer, Anton E Kunst, Regina M van der Meer, Hein de Vries, Boukje M van Gelder, and Marc C Willemsen. Trends in socioeconomic inequalities in smoking prevalence, consumption, initiation, and cessation between 2001 and 2008 in the netherlands. findings from a national population survey. *BMC public health*, 12(1):1–9, 2012.

Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45 (2):167–256, 2003.

Marie Ng, Michael K Freeman, Thomas D Fleming, Margaret Robinson, Laura Dwyer-Lindgren, Blake Thomson, Alexandra Wollum, Ella Sanman, Sarah Wulf, Alan D Lopez, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. *Jama*, 311(2):183–192, 2014.

Javed Khan OBE. The khan review making smoking obsolete. *Office for Health Improvement and Disparities*, 25, 2022.

Office of National Statistics. Do smoking rates vary between more and less advantaged areas? 2014.

Office for National Statistics. Adult smoking habits in the UK: 2019. *Statistical bulletin*, pages 1–15, 2019. URL https://backup.ons.gov.uk/wp-content/uploads/sites/3/2019/07/Adult-smoking-habits-in-the-UK-2018.pdf%0Ahttps://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2019.

Office of United States Public Health Service and others. Smoking cessation: A report of the surgeon general [internet]. Technical report, 2020.

Jennifer L O'Loughlin, Erika N Dugas, Erin K O'Loughlin, Igor Karp, and Marie-Pierre Sylvestre. Incidence and determinants of cigarette smoking initiation in young adults. *Journal of Adolescent Health*, 54(1):26–32, 2014.

Magdalena Opazo Breton, Duncan Gillespie, Robert Pryce, Ilze Bogdanovica, Colin Angus, Monica Hernandez Alava, Alan Brennan, and John Britton. Understanding long-term trends in smoking in england, 1972–2019: an age–period–cohort approach. *Addiction*, 117(5):1392–1403, 2022.

World Health Organization. *WHO technical manual on tobacco tax administration*. World Health Organization, 2010.

World Health Organization et al. *MPOWER: a policy package to reverse the tobacco epidemic*. World Health Organization, 2008.

Mark G Orr, George A Kaplan, and Sandro Galea. Neighbourhood food, physical activity, and educational environments and black/white disparities in obesity: a complex systems simulation analysis. *J Epidemiol Community Health*, 70(9):862–867, 2016.

John Oyston. A fresh approach to tobacco control: raising the minimum legal age for access. *CMAJ*, 189(8):E293–E294, 2017.

New Zealand Parliament. Smokefree environments and regulated products (smoked tobacco) amendment bill. *Wellington: New Zealand Parliament*, 2022.

John P Pierce, Janet M Distefan, Robert M Kaplan, and Elizabeth A Gilpin. The role of curiosity in smoking initiation. *Addictive behaviors*, 30(4):685–696, 2005.

Kirstin Pirie, Richard Peto, Gillian K Reeves, Jane Green, and Valerie Beral. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the uk. *The Lancet*, 381(9861):133–141, 2013.

Charlotta Pisinger, Mette Aadahl, Ulla Toft, and Torben Jørgensen. Motives to quit smoking and reasons to relapse differ by socioeconomic status. *Preventive medicine*, 52(1):48–52, 2011.

Steven B Pokorny, Leonard A Jason, and Michael E Schoeny. The relation of retail tobacco availability to initiation and continued smoking. *Journal of Clinical Child and Adolescent Psychology*, 32(2):193–204, 2003.

Adarsh Prabhakaran, Valerio Restocchi, and Benjamin D Goddard. Networks for smoking dynamics. In *The 11th International Conference on Complex Networks and their Applications 2022*, 2022.

Adarsh Prabhakaran, Valerio Restocchi, and Benjamin D Goddard. Improving tobacco social contagion models using agent-based simulations on networks. *Applied Network Science*, 8(1):1–21, 2023.

Hazhir Rahmandad and John Sterman. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management science*, 54(5):998–1014, 2008.

Marissa B Reitsma, Nancy Fullman, Marie Ng, Joseph S Salama, Amanuel Abajobir, Kalkidan Hassen Abate, Cristiana Abbafati, Semaw Ferede Abera, Biju Abraham, Gebre Yitayih Abyu, et al. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the global burden of disease study 2015. *The Lancet*, 389(10082):1885–1906, 2017.

Kurt M Ribisl, Douglas A Luke, Doneisha L Bohannon, Amy A Sorg, and Sarah Moreland-Russell. Reducing disparities in tobacco retailer density by banning tobacco product sales near schools. *Nicotine & Tobacco Research*, 19(2):239–244, 2017.

Kristen E Riley, Michael R Ulrich, Heidi A Hamann, and Jamie S Ostroff. Decreasing smoking but increasing stigma? anti-tobacco campaigns, public health, and cancer care. *AMA Journal of Ethics*, 19(5):475, 2017.

Mark Robinson and Garth Reid. Population cigarette consumption in great britain: novel insights using retail sales data. *BMC Public Health*, 17(1):1–7, 2017.

Shyanika W Rose, Allison E Myers, Heather D'Angelo, and Kurt M Ribisl. Peer reviewed: Retailer adherence to family smoking prevention and tobacco control act, north carolina, 2011. *Preventing chronic disease*, 10, 2013.

Harry Rutter, Natalie Savona, Ketevan Glonti, Jo Bibby, Steven Cummins, Diane T Finegood, Felix Greaves, Laura Harper, Penelope Hawe, Laurence Moore, et al. The need for a complex systems model of evidence for public health. *The lancet*, 390(10112):2602–2604, 2017.

Antti J Saari, Jukka Kentala, and Kari J Mattila. The smoking habit of a close friend or family member—how deep is the impact? a cross-sectional study. *BMJ open*, 4 (2):e003218, 2014.

James D Sargent, Joy Gabrielli, Alan Budney, Samir Soneji, and Thomas A Wills. Adolescent smoking experimentation as a predictor of daily cigarette smoking. *Drug and alcohol dependence*, 175:55–59, 2017.

Francesca Scalici and Peter J Schulz. Parents' and peers' normative influence on adolescents' smoking: results from a swiss-italian sample of middle schools students. *Substance abuse treatment, prevention, and policy*, 12:1–9, 2017.

David R Schaefer, Steven A Haas, and Nicholas J Bishop. A dynamic model of us adolescents' smoking and friendship networks. *American journal of public health*, 102(6):e12–e18, 2012.

David R Schaefer, Jimi Adams, and Steven A Haas. Social networks and smoking: exploring the effects of peer influence and smoker popularity through simulations. *Health Education & Behavior*, 40(1_suppl):24S–32S, 2013.

Shari Kessel Schneider, Stephen L Buka, Kim Dash, Jonathan P Winickoff, and Lydia O'Donnell. Community reductions in youth smoking after raising the minimum tobacco sales age to 21. *Tobacco control*, 25(3):355–359, 2016.

Sven Schneider, Johannes Gruber, Shelby Yamamoto, and Christian Weidmann. What happens after the implementation of electronic locking devices for adolescents at cigarette vending machines? a natural longitudinal experiment from 2005 to 2009 in germany. *Nicotine & Tobacco Research*, 13(8):732–740, 2011.

ScotCen Social Research. Scottish health survey, 2021, 2023.

Scottish Executive. The smoking, health and social care (scotland) act 2005. *Smoking: prohibition and control*, 2005.

Scottish Government. Raising scotland's tobacco free generation: our tobacco control action plan, 2018.

Scottish Government. Scottish index of multiple deprivation 2020. *Office of the Chief Statistician, Scottish Executive*, 10, 2020.

Scottish Government. Households and dwellings in scotland, 2021, 2021a.

Scottish Government. Population estimates by scottish index of multiple deprivation (simd), 2021b.

Scottish Government. Scottish school roll and locations, 2022a.

Scottish Government. Business stocks and sites by urban rural classification, 2022b.

Swarnali Sharma and GP Samanta. Analysis of a drinking epidemic model. *International Journal of Dynamics and Control*, 3(3):288–305, 2015.

Oluwaseun Sharomi and Abba B Gumel. Curtailing smoking dynamics: a mathematical modeling approach. *Applied Mathematics and Computation*, 195(2):475–499, 2008.

Huiyoung Shin. Social contagion of academic behavior: Comparing social networks of close friends and admired peers. *Plos one*, 17(3):e0265385, 2022.

Niamh K Shortt, Catherine Tisch, Jamie Pearce, Richard Mitchell, Elizabeth A Richardson, Sarah Hill, and Jeff Collin. A cross-sectional analysis of the relationship between tobacco and alcohol outlet density and neighbourhood deprivation. *BMC public health*, 15(1):1–9, 2015.

Niamh K Shortt, Helena Tunstall, Richard Mitchell, Emma Coombes, Andy Jones, Garth Reid, and Jamie Pearce. Using point-of-sale data to examine tobacco pricing across neighbourhoods in scotland. *Tobacco Control*, 30(2):168–176, 2021.

NK Shortt, Catherine Tisch, Jamie Pearce, EA Richardson, and Richard Mitchell. The density of tobacco retailers in home and school environments and relationship with adolescent smoking behaviours in scotland. *Tobacco control*, 25(1):75–82, 2016.

Crystal R Smit, Rebecca NH De Leeuw, Kirsten E Bevelander, William J Burk, Laura Buijs, Thabo J Van Woudenberg, and Moniek Buijzen. Promoting water consumption among children: A three-arm cluster randomised controlled trial testing a social network intervention. *Public health nutrition*, 24(8):2324–2336, 2021.

ChangeLab Solutions. Tobacco retailer density: Place-based strategies to advance health and equity. *Oakland, CA: ChangeLab Solutions*, 2019.

Lindsay F Stead, Priya Koilpillai, Thomas R Fanshawe, and Tim Lancaster. Combined pharmacotherapy and behavioural interventions for smoking cessation. *Cochrane database of systematic reviews*, (3), 2016.

Laurence Steinberg and Kathryn C Monahan. Age differences in resistance to peer influence. *Developmental psychology*, 43(6):1531, 2007.

Strathclyde Partnership for Transport. Glasgow and strathclyde transport act scoping study : Urban bus transport mobility comparison, 2022.

Silvia Stringhini, Séverine Sabia, Martin Shipley, Eric Brunner, Hermann Nabi, Mika Kivimaki, and Archana Singh-Manoux. Association of socioeconomic position with health behaviors and mortality. *Jama*, 303(12):1159–1166, 2010.

Donald H Taylor Jr, Vic Hasselblad, S Jane Henley, Michael J Thun, and Frank A Sloan. Benefits of smoking cessation for longevity. *American journal of public health*, 92(6):990–996, 2002.

Hadi Tehrani, Abdolhalim Rajabi, Mousa Ghelichi-Ghojogh, Mahbobeh Nejatian, and Alireza Jafari. The prevalence of electronic cigarettes vaping globally: a systematic review and meta-analysis. *Archives of Public Health*, 80(1):1–15, 2022.

The Lancet Public Health. Making progress in the fight against tobacco. *The Lancet Public Health*, 4(12):e592, 2019. ISSN 2468-2667. doi: https://doi.org/10.1016/S2468-2667(19)30228-2. URL https://www.sciencedirect.com/science/article/pii/S2468266719302282.

Stefan Thurner, Peter Klimek, and Rudolf Hanel. A network-based explanation of why most covid-19 infection curves are linear. *Proceedings of the National Academy of Sciences*, 117(37):22684–22689, 2020.

Sigrid A Troelstra, Anton E Kunst, and Janneke Harting. “like you are fooling yourself”: how the “stoptober” temporary abstinence campaign supports dutch smokers attempting to quit. *BMC Public Health*, 19(1):1–10, 2019.

United States - Centers for Disease Control and Prevention (CDC). National youth tobacco survey, 2022.

US Department of Health and Human Services and others. The health consequences of smoking-50 years of progress: a report of the surgeon general, 2014.

US Department of Health and Human Services and others. Smoking cessation: a report of the surgeon general. *Atlanta: US Department of Health and Human Services*, 2020.

Ari Väänänen, Anne Kouvonen, Mika Kivimäki, Jaana Pentti, and Jussi Vahtera. Social support, network heterogeneity, and smoking behavior in women: the 10-town study. *American Journal of Health Promotion*, 22(4):246–255, 2008.

Thomas W Valente. Network interventions. *science*, 337(6090):49–53, 2012.

Donna M Vallone, Jane A Allen, and Haijun Xiao. Is socioeconomic status associated with awareness of and receptivity to the truth campaign? *Drug and alcohol dependence*, 104:S115–S120, 2009.

Martijn Van Hasselt, Judy Kruger, Beth Han, Ralph S Caraballo, Michael A Penne, Brett Loomis, and Joseph C Gfroerer. The relation between tobacco taxes and youth and young adult smoking: what happened following the 2009 us federal tax increase on cigarettes? *Addictive behaviors*, 45:104–109, 2015.

Jeroen van Meijgaard and Jonathan E Fielding. Estimating benefits of past, current, and future reductions in smoking rates using a comprehensive model with competing causes of death. *Prev Chronic Disease*, 2012.

Eleni Vangeli and Robert West. Sociodemographic differences in triggers to quit smoking: findings from a national survey. *Tobacco control*, 17(6):410–415, 2008.

Alexey Voinov, Karen Jenni, Steven Gray, Nagesh Kolagani, Pierre D Glynn, Pierre Bommel, Christina Prell, Moira Zellner, Michael Paolisso, Rebecca Jordan, et al. Tools and methods in participatory modeling: Selecting the right tool for the job. *Environmental Modelling & Software*, 109:232–255, 2018.

Robert B Wallace, Amy B Geller, V Ayano Ogawa, et al. *Assessing the use of agent-based models for tobacco regulation*. National Academies Press Washington, DC, 2015.

Teresa W Wang, Linda J Neff, Eunice Park-Lee, Chunfeng Ren, Karen A Cullen, and Brian A King. E-cigarette use among middle and high school students—united states, 2020. *Morbidity and Mortality Weekly Report*, 69(37):1310, 2020.

Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’networks. *nature*, 393(6684):440–442, 1998.

Warren Weaver. Science and complexity. *American scientist*, 36(4):536–544, 1948.

Emma White and Catherine Comiskey. Heroin epidemics, treatment and ode modelling. *Mathematical biosciences*, 208(1):312–324, 2007.

Anna L Wilkinson, Michelle M Scollo, Melanie A Wakefield, Matthew J Spittal, Frank J Chaloupka, and Sarah J Durkin. Smoking prevalence following tobacco tax increases in australia between 2001 and 2017: an interrupted time-series analysis. *The Lancet Public Health*, 4(12):e618–e627, 2019.

Jonathan P Winickoff, Mark Gottlieb, and Michelle M Mello. Tobacco 21—an idea whose time has come. *New England Journal of Medicine*, 370(4):295–297, 2014.

World Health Organization. Tobacco. URL https://www.who.int/news-room/fact-sheets/detail/tobacco.

World Health Organization et al. Who report on the global tobacco epidemic, 2023: protect people from tobacco smoke. 2023.

World Health Organization and others. *WHO framework convention on tobacco control*. World Health Organization, 2003.

World Health Organization and others. Who framework convention on tobacco control. Technical report, WHO Regional Office for South-East Asia, 2004.

World Health Organization and others. Who global report on trends in prevalence of tobacco use 2000-2025. Technical report, 2019.

World Health Organization and others. *WHO global report on trends in prevalence of tobacco use 2000-2025*. World Health Organization, 2021.

Angela Difeng Wu, Min Gao, Paul Aveyard, and Gemma Taylor. Smoking cessation and changes in anxiety and depression in adults with and without psychiatric disorders. *JAMA Network Open*, 6(5):e2316111–e2316111, 2023.

David Young, Ron Borland, and Ken Coghill. Changing the tobacco use management system: blending systems thinking with actor–network theory. *Review of Policy Research*, 29(2):251–279, 2012.

Gul Zaman. Optimal campaign in the smoking dynamics. *Computational and Mathematical Methods in Medicine*, 2011, 2011a.

Gul Zaman. Qualitative behavior of giving up smoking models. *Bulletin of the Malaysian Mathematical Sciences Society*, 34(2):403–415, 2011b.

Gul Zaman, Young Han Kang, and Il Hyo Jung. Dynamics of a smoking model with smoking death rate. *Applicationes Mathematicae*, 44:281–295, 2017.

Bin Zhou, Sen Pei, Lev Muchnik, Xiangyi Meng, Xiaoke Xu, Alon Sela, Shlomo Havlin, and H Eugene Stanley. Realistic modelling of information spread using peer-to-peer diffusion patterns. *Nature Human Behaviour*, 4(11):1198–1207, 2020.