MODELLING THE EFFECTIVENESS AND EQUITY OF PRIMARY PREVENTION POLICIES IN ENGLAND

Christodoulos Kypridemos
Doctor of Philosophy (PhD)
October 2016

MODELLING THE EFFECTIVENESS AND EQUITY OF PRIMARY PREVENTION POLICIES IN ENGLAND

A Stochastic Dynamic Microsimulation for the Joint Prevention of Non Communicable Diseases

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Christodoulos Kypridemos

Doctor of Philosophy (PhD)
Department of Public Health and Policy
Institute of Psychology Health and Society
University of Liverpool

October 2016

Christodoulos Kypridemos. Modelling the effectiveness and equity of primary prevention policies in England: a stochastic dynamic microsimulation for the joint prevention of non communicable diseases, Doctor of Philosophy (PhD), © October 2016

SUPERVISORS:
Martin O'Flaherty
Simon Capewell
Iain Buchan

LOCATION:
Liverpool, UK

TIME FRAME:
October 2013 to October 2016

Dedicated to my beloved wife Iliana
who tolerates my presence and my absences

ABSTRACT

CHRISTODOULOS KYPRIDEMOS. MODELLING THE EFFECTIVENESS AND EQUITY OF PRIMARY PREVENTION POLICIES IN ENGLAND

INTRODUCTION: Cardiovascular disease and cancers are the main causes of premature death and disability in England. This thesis uses a microsimulation modelling methodology to examine and quantify the effectiveness and equity of existing primary prevention policies and feasible alternatives.

METHODS: I created and validated IMPACT$_{NCD}$, a dynamic stochastic microsimulation model from first epidemiological principles, to simulate the life course of synthetic individuals under counterfactual scenarios. First, I used the model to quantify the contribution of statins to the observed decline in total cholesterol in England. Then, I examined a national screening programme known as 'NHS Health Checks'. Afterwards, I estimated the effectiveness and equity of the national salt reduction strategy. Finally, I studied two proposed policies for the tobacco 'endgame'; a total sales ban, and a sales ban restricted to those born in or after 2000.

RESULTS: The model suggested that statins contributed only about a third of the observed total cholesterol decline in England since 1991-92. Their impact on reducing socioeconomic inequalities in total cholesterol was generally positive, contrary to what was anticipated. NHS Health Checks may prevent or postpone about 19 000 cases of cardiovascular disease by 2030; however, population wide structural policies could be three times more effective and generally more equitable. IMPACT$_{NCD}$ estimated that the national salt reduction strategy may have prevented or postponed about 52 000 cases of cardiovascular disease and 5000 cases of gastric cancer since 2003. Additional legislative policies from 2016 onwards could further prevent or postpone approximately 20 000 more cases by 2030, while reducing inequalities. Finally, a total ban on sales of tobacco products could prevent or postpone about 90 000 cases of cardiovascular disease, 79 000 cases of lung cancer, and tremendously reduce health inequalities by 2045. The age restricted ban could have small benefits overall within the simulation horizon.

CONCLUSIONS: Increasing the structural elements of existing policies or complementing them with new structural policies might maximise their effectiveness and equity. Simulation modelling is valuable for the evaluation of existing policies and the design of new fit for purpose policies that will take into account the complex nature and dynamics of the populations.

PUBLICATIONS

The core IMPACT$_{NCD}$ engine has been published in: Kypridemos C et al. Cardiovascular screening to reduce the burden from cardiovascular disease: microsimulation study to quantify policy options. BMJ 2016;353:i2793.

Some ideas and figures have appeared previously in the following peer reviewed publications and abstracts:

FULL PAPERS

1. Kypridemos C, O'Flaherty M and Capewell S. Fruit and vegetable consumption and non-communicable disease: time to update the '5 a day' message? J Epidemiol Community Health 2014;68:799–800. DOI: 10.1136/jech-2014-203981.

2. Kypridemos C, Bandosz P, Hickey GL et al. Quantifying the contribution of statins to the decline in population mean cholesterol by socioeconomic group in England 1991 - 2012: a modelling study. PLOS One 2015;10:e0123112. DOI: 10.1371/journal.pone.0123112.

3. Bandosz P, O'Flaherty M, Rutkowski M et al. A victory for statins or a defeat for diet policies? Cholesterol falls in Poland in the past decade: a modeling study. Int J Cardiol 2015;185:313–319. DOI: 10.1016/j.ijcard.2015.03.079.

4. Allen K, Kypridemos C, Hyseni L et al. The effects of maximising the UK's tobacco control score on inequalities in smoking prevalence and premature coronary heart disease mortality: a modelling study. BMC Public Health 2016;16:292. DOI: 10.1186/s12889-016-2962-8.

5. Pearson-Stuttard J, Guzman Castillo M, Penalvo JL et al. Modelling future cardiovascular disease mortality in the United States: national trends and racial and ethnic disparities. Circulation 2016;133:967–78. DOI: 10.1161/CIRCULATIONAHA.115.019904.

ABSTRACTS

1. Bandosz P, O'Flaherty M, Capewell S et al. PP05 Are recent population declines in total cholesterol levels in Poland attributable to statin medications or dietary changes? J Epidemiol Community Health Suppl 1 2014;68:A49. DOI: 10.1136/jech-2014-204726.103.

2. Castillo MG, O'Flaherty M, Couch P et al. OP71 Comparing coronary mortality reductions by shifting the population blood pressure distribution versus improved management of hypertensive patients: modelling study. J Epidemiol Community Health Suppl 1 2014;68:A35–A36. DOI: 10.1136/jech-2014-204726.73.

3. Allen K, Kypridemos C, Hyseni L et al. OP11 The effects of maximising the UK's tobacco control score on inequalities in smoking prevalence and premature coronary heart disease mortality: a modelling study. J Epidemiol Community Health Suppl 1 2015;69:A13. DOI: 10.1136/jech-2015-206256.11.

4. Kypridemos C, Allen K, Bandosz P et al. OP10 Quantifying national policy options for equitable primary cardiovascular disease prevention in England up to 2025: a micro-simulation study. J Epidemiol Community Health Suppl 1 2015;69:A13. DOI: 10.1136/jech-2015-206256.10.

5. Kypridemos C, Bandosz P, Hickey G et al. OP04 Quantifying the contribution of statins to the decline in population mean cholesterol by socioeconomic group in England 1991–2012: a modelling study. J Epidemiol Community Health Suppl 1 2015;69:A10–A10. DOI: 10.1136/jech-2015-206256.4.

6. Kypridemos C, Guzman-Castillo M, Hyseni L et al. An opportunity to reduce the burdens and inequalities in cardiovascular disease and gastric cancer caused by dietary salt: IMPACT$_{NCD}$ microsimulation study. Lancet 2015. Public Health Science November 2015 386, Supplement 2:S47. DOI: 10.1016/S0140-6736(15)00885-5.

7. Hyseni L, Elliot-Green A, Lloyd-Williams F et al. Abstract P154: systematic review of dietary salt reduction policies: evidence for an effectiveness hierarchy? Circulation Suppl 1 2016;133:AP154. Available at: http://circ.ahajournals.org/content/133/Suppl_1/AP154 (accessed on 4th Apr. 2016).

*"You can analyse the past, but you
have to design the future."* — *Edward De Bono*

ACKNOWLEDGEMENTS

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AMI   acute myocardial infarction

BMI   body mass index

CHD   coronary heart disease

CI   confidence interval

CVD   cardiovascular disease

DALY   disability-adjusted life year

GB   Great Britain

HSE   Health Survey for England

ICD10   International Classification of Diseases, version 10

IQR   interquartile range

LDL   low density lipoprotein

NCD   non-communicable disease

NDNS   National Diet and Nutrition Survey

NHS   National Health Service

NICE   National Institute for Health and Care Excellence

ONS   Office for National Statistics

PA   physical activity
PAF   population attributable fraction

QIMD   quintile groups of Index of Multiple Deprivation

RII   relative index of inequality

SBP   systolic blood pressure
SD   standard deviation
SE   standard error
SII   slope index of inequality

UI   uncertainty interval
UK   United Kingdom
US   United States

WHO   World Health Organisation

Part I

BACKGROUND AND MODELLING FRAMEWORK

GENERAL INTRODUCTION

The 20th century was an exciting time for public health worldwide and in the United Kingdom (UK) specifically. Improvements in living conditions, better hygiene, universal immunisation programmes, and effective treatments led to the control of infectious diseases in high- and middle-income countries. Infant and child mortality declined rapidly and in the UK, the National Health Service (NHS) was established to provide universal healthcare free at the point of need. Over the last century, life expectancy at birth increased by almost 30 years in the UK and reflects the public health achievements of the era.

Despite the numerous successes in the previous century, a substantial proportion of deaths and disability worldwide remains preventable and potentially avoidable. For high- and middle-income countries cardiovascular disease (CVD), cancers, and other non-communicable diseases (NCDs) cause the highest burden of avoidable morbidity and mortality. Many of today's public health challenges have their roots in the 20th century. The emergence of NCDs, the tobacco, and the obesity epidemics are a few striking examples. Similarly to the infectious disease epidemics of previous centuries, these new challenges are immensely preventable. Causes, challenges, and solutions have only been evolved. Infectious agents have now been overtaken by contagions of tobacco and food industry tactics; the need for food hygiene has been replaced by the need for healthy and sustainable food supply systems; and the need for smog free cities has evolved into the more wide ranging need for a cleaner environment. As epidemics in previous centuries shared poverty, poor hygiene, and unhealthy living conditions of the population as common determinants; current NCD epidemics share tobacco, unhealthy diet, and sedentary lifestyle.

Inequalities in health have also evolved over time. A lot has changed since Edwin Chadwick and the Public Health Act in 1848. Undeniably, living conditions and health have been improved in absolute terms in the course of time. However, dramatic differences in health related outcomes and health remain between countries and more worryingly within countries. In the UK, striking differences exist in life and healthy life expectancy of sub-populations, powerfully reflecting inequalities in the burden of NCDs and overall health throughout the life course. Reassuringly, tackling inequalities has recently become a priority for policy makers and the public health community in the UK.

It is well accepted now that a large proportion of health inequalities in the UK and elsewhere can be attributed to existing socioeconomic inequalities. Nevertheless, tackling socioeconomic inequalities in health can be achieved even without radical changes in the established socioeconomic conditions. This is possible by identifying and disassembling the mechanisms that generate health inequalities from socioeconomic inequalities and by designing and implementing health policies that can reverse the association between

low socioeconomic status and poor health. Despite current theoretical evidence regarding the type of policies that could achieve this, it appears that their implementation is the exception rather than the rule. In fact, some of the existing health promotion policies may have even increased socioeconomic inequalities in health.

Hence, policy makers need to maximise effectiveness and equity of current and future public health policies and to bridge the gap between the evidence base and the formulation of such policies. Simulation modelling can provide policy makers with a tool that integrates all the relevant information from multiple sources and the complex dynamics within the population in order to design better policies. Modelling can provide a platform where potential future policies can be evaluated and rejected or improved before implementation. Furthermore, modelling can be used to analyse the effectiveness and equity of existing policies, explain their efficiencies and inefficiencies, and complement traditional evaluation methods.

This thesis aims to provide further evidence regarding the type of primary prevention policies that are likely to be both effective and decrease socioeconomic health inequalities. It also aspires to be a proof of concept that simulation modelling is mature enough to inform public health policy and be used as a decision support tool from policy makers, planners, and practitioners.

THESIS STRUCTURE

Elements in this thesis span from epidemiology and public health policy to computational statistics and software engineering. I hope the readers will appreciate that this is the very nature of simulation modelling. The modeller has to understand the phenomenon and its dynamics, reduce it to its main components, and finally recreate it computationally.

CHAPTER 1   All the background information that is necessary to conceptualise the aims and objectives of the thesis. It includes an overview of the epidemiology of CVD, lung and gastric cancers, and their associate modifiable risk factors; a conceptual framework of the social determinants of health and how socioeconomic inequalities in health are generated; a typology of primary preventive interventions; and finally the aims and objectives of this thesis.

CHAPTER 2   A detailed description of IMPACT$_{NCD}$, the model that was used in this thesis. The concepts are presented in a non-technical manner and from an epidemiological perspective.

CHAPTER 3   An extensive validation of IMPACT$_{NCD}$ main components. Face, internal, cross, and external validity are assessed.

CHAPTER 4   The first chapter of the results. The contribution of statins to the observed serum total cholesterol decline in England since 1991 is assessed and quantified.

CHAPTER 5   The potential impact and equity of a national screening programme for primary CVD prevention known as 'NHS Health Checks' is assessed and quantified. It is then compared with feasible alternative policies.

CHAPTER 6   The potential impact and equity of the national salt reduction strategy is assessed and quantified. Current strategy is then compared with a stricter legislative strategy.

CHAPTER 7   Two proposed policies for the tobacco 'endgame'; a total sales ban and the 'tobacco free millennial generation' proposal are simulated. The focus here is on the dynamics of the effectiveness and equity of the two policies.

CHAPTER 8   The general discussion. The emerging overarching themes are discussed. Implications for public health policy makers, planners, and practitioners are discussed. Limitations, future plans, and reflections are offered.

CHAPTER 9   Conclusions.

1.1 AN OVERVIEW OF CARDIOVASCULAR DISEASE EPIDEMIOLOGY

In this section, I will briefly describe the current burden of CVD globally and in the UK specifically. CVD incidence and mortality rates have been declining for at least four decades; however, ageing of the population and the recent obesity epidemic partly counteract this decline. Despite the overall decline, relative socioeconomic inequalities persist and may have even increased.

1.1.1 *Definitions*

CVD includes a spectrum of diseases related to the heart, vascular diseases of the brain, and diseases of blood vessels. Many of these seemingly unrelated diseases share a common pathophysiologic mechanism known as atherothrombosis. Atherothrombosis is the outcome of the deposition of fatty material on the inner wall of the arteries, which leads to their thickening and stiffening (atherosclerosis). The disruption of the atherosclerotic lesions promotes blood clotting and the formation of thrombus (atherothrombosis). Thus, CVD can be classified into two broad categories: CVD related to atherosclerosis and all other CVD.[1] The first category includes coronary heart disease (CHD), stroke (cerebrovascular disease), and diseases of the aorta and arteries, including hypertension and peripheral arterial disease. This category poses a greater interest to public health and clinicians for two reasons; 1. it is preventable and 2. its burden on the population worldwide is enormously high.

1.1.2 *Worldwide mortality and trends*

CVD is the leading cause of death worldwide. Of the 54.9 million deaths that have been estimated that occurred in 2013, 17.3 million have been attributed to CVD; a 41 % increase since 1990. Of those 17.3 million CVD deaths, 8.1 million were attributable to CHD and 6.4 million to stroke. The main reason for the observed increase in CVD deaths was population ageing. Over the same period the age-standardised CVD mortality rate declined from 375.5 to 293.2 per 100 000, a 22 % drop.[2, 3]

1.1.3 *Burden in England and the United Kingdom*

Throughout this thesis I have consistently tried to use data specifically for England. When data were not available, I have used aggregated data for the UK or Great Britain (GB). Since the population of England is considerably larger than the population in all other countries in the UK, aggregated UK data are mostly representative for England.

1.1.3.1 *Mortality*

For decades, CVD used to be the leading cause of death in the UK. Since 2012, deaths from cancer have surpassed those from CVD, leaving CVD as the second leading cause

of mortality. In 2014, CVD was responsible for 154 639 deaths, approximately 27 % of all deaths recorded in the UK that year. Within the CVD spectrum, CHD was responsible for 69 163 deaths and stroke for 39 282 deaths over the same year. From the four countries in the UK, England has the lowest age-standardised mortality and Scotland the highest. Specifically for England in 2014, 126 682 deaths were recorded due to CVD, from which 56 364 were attributed to CHD and 31 787 were attributed to stroke in the same year. In general, CVD mortality is higher among men and increases almost exponentially with age.[4, 5]

1.1.3.2 *Incidence*

The true number of new CVD cases is largely unknown. However, the annual records of inpatient episodes from NHS hospitals were used by the Cardiovascular Disease Statistics 2015 report to crudely approximate the true incidence.[5] According to the report, there were 793 952 inpatient episodes of CVD among men in England in 2013/14. Of those, 264 934 were due to CHD and 97 593 were due to stroke. The inpatient episodes of CVD among women in the same period were 607 280 of which 136 073 due to CHD and 99 763 due to stroke. The two main limitations of these numbers are that patients having recurrent admissions with the same diagnosis were counted multiple times; while the potentially large number of patients that died before reaching medical care were not recorded at all.

To overcome the second limitation Smolina *et al.* combined data from the records of inpatient episodes with reported mortality from the Office for National Statistics (ONS).[6, 7] Smolina *et al.* estimated that the incidence of acute myocardial infarction (AMI)[1] for England in 2010 was 82 252 cases; 63 864 from hospital admissions records and 18 388 from sudden AMI deaths. They also estimated the incidence rate for all ages over 29 years old to be 154 cases per 100 000 for men and 66 cases per 100 000 for women. Figure 1.1 on the facing page depicts the age-standardised incidence rates by age group and sex as reported by Smolina *et al.*[6]

The lack of an accurate, unbiased estimate is similar for stroke incidence. It is worth noting that unlike CHD, stroke incidence appears similar for both sexes and is possibly slightly higher among women. The most detailed and likely less biased source of information, regarding stroke incidence, appears to be the 'Oxford vascular study'.[8, 9] However, this is largely becoming outdated because it was conducted more than 10 years ago. It may also not be nationally representative, as the researchers collected data only from the relatively affluent area of Oxfordshire. Therefore, uncertainty regarding CVD incidence remains. I will describe later in the methods chapter how I addressed this challenge.

1.1.3.3 *Prevalence*

As with CVD incidence, accurate estimation of CVD prevalence in the population is very difficult. The Cardiovascular Disease Statistics 2015 report used the Quality and Outcomes

---

1 Acute myocardial infarction is one of the two clinical presentation of CHD. Angina pectoris is the second one.
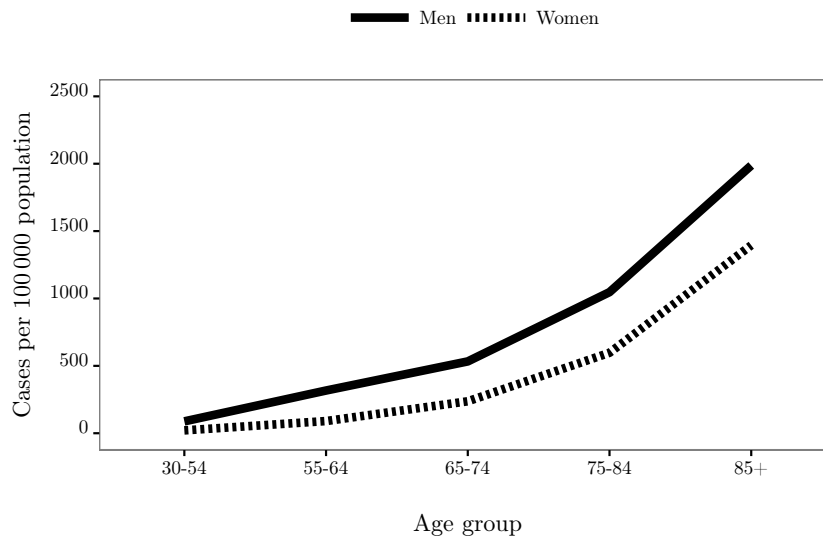
Figure 1.1: Incidence rate of acute myocardial infarction by age group and sex in England, 2010. Incident cases include all mortalities and hospital admissions for acute myocardial infarction (International Classification of Diseases, version 10 I21-22) with no previous hospital admission for the same condition in the previous 30 days. Incident cases potentially include misdiagnoses and further investigation of earlier acute myocardial infarctions. Directly age-standardised to the European Standard Population.[10] Data source: Smolina *et al.*[6]

Framework[2] to estimate CVD prevalence.[5] According to the report, there are approximately 1.86 million patients in the CHD registry and 0.97 million patients in the stroke registry in England; some 3.3 % and 1.7 % of the population, respectively.

Another source of information regarding CVD prevalence is the Health Survey for England (HSE). HSE is a nationally representative health survey of the community dwelling population in England (please refer to section 2.3.1 on page 46 for a more detailed description). In HSE2011, among adults aged 16 and over, 13.9 %of men and 13.4 % of women self-reported that they had been diagnosed with CVD. Self-reported prevalence increased with age: from 3 % of men and 5 % of women aged 16 to 24, to 54 % of men and 31 % of women aged 85 and over.[11, chapter 2]

1.1.4 Trends in England and the United Kingdom

CVD burden in the UK has evolved over time producing from epidemiological perspective, very interesting patterns.

---

2 The Quality and Outcomes Framework became part of general practice contracts in 2004. General practices are financially rewarded for keeping records of the number of registered patients who have been diagnosed with certain conditions.
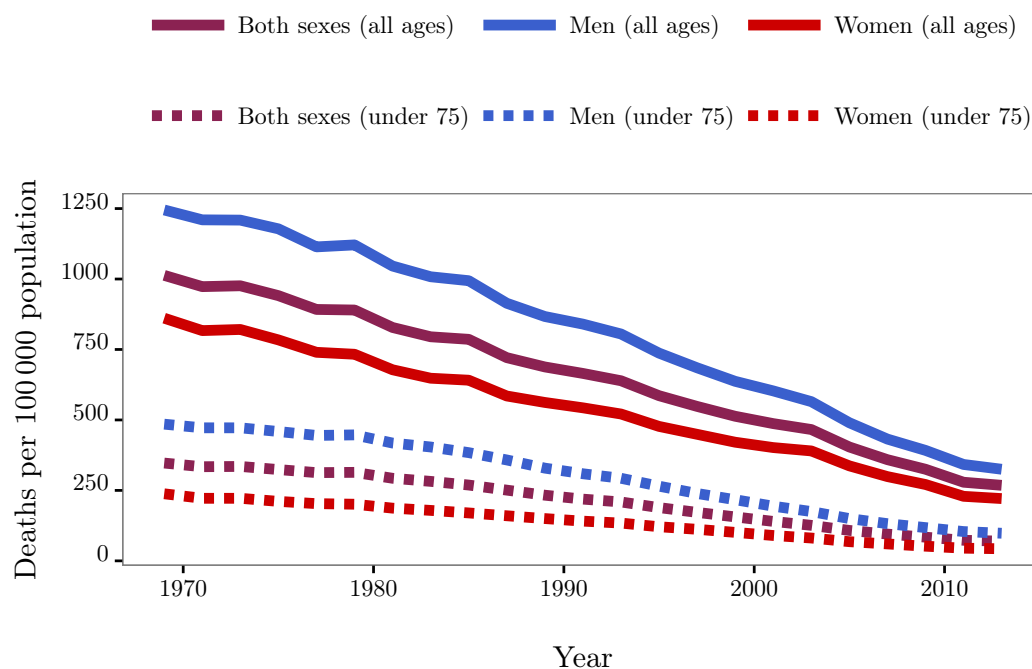
Figure 1.2: Age-standardised cardiovascular disease mortality rate trends in England, 1969 – 2013 (all ages and under 75). Directly age-standardised to the European Standard Population 2013.[10] Data source: Cardiovascular Disease Statistics 2015 report.[5]

1.1.4.1   *Mortality trends*

In contrast to the upward trend in absolute number of CVD deaths worldwide (section 1.1.3.1 on page 7), mortality in the UK has been declining both in absolute numbers and in age-standardised mortality rates in recent years.[5] This is similar to the mortality trends in other high-income countries.[3] Age-standardised CVD mortality rates have fallen by 74 % in England since 1969. The sharp decrease counteracted the effect of population ageing over the same period and it was more pronounced in older ages and among men (figure 1.2).

Declining CVD mortality trend is the combining outcome of declining CVD incidence (primary prevention) and better survival for those already diseased (secondary and tertiary prevention). Specifically, for CHD the original IMPACT model suggested that about 80 % of the observed mortality decline between 1980 and 2000 could be attributed to primary prevention and only about 20 % to better treatments.[12] Smolina *et al.* later estimated that for the more recent period 2002 to 2010, the contribution of treatments almost counterbalanced the contribution of primary prevention.[6]

1.1.4.2   *Incidence trends*

Despite the difficulties in measuring the absolute number of new cases every year (section 1.1.3.2 on page 8), incidence trends are more easily monitored when the same method

is applied consistently every year. Hence, following the Cardiovascular Disease Statistics 2015 report method, which used the annual records of inpatient episodes from NHS hospitals, overall CVD incidence has been increasing since 2005 for both sexes.[5] In contrast, CHD cases have been mildly decreasing for both sexes despite population ageing. Finally, stroke inpatient episodes increased between 2006 and 2009 and have remained steady since.

Smolina *et al.* combined data from the records of inpatient episodes with reported mortality from ONS and they estimated incidence rates for AMI.[6] According to their findings, the AMI annual age-standardised incidence rate for men fell from approximately 230 to 154 per 100 000 population between 2002 and 2010. Similarly, the annual age-standardised incidence rate for women fell from about 95.4 to 66 per 100 000 over the same period. AMI incidence rate reductions were observed across all ages, although these were smaller for ages under 55 and over 85.

Specifically for stroke, Lee *et al.* analysed data from the General Practice Research Database[3], a database from about 500 practices in the UK that covers a population of more than three million. They reported a 30 % reduction in first stroke incidence between 1999 and 2008. The reduction appeared to be larger in older age groups.[13]

1.1.4.3 *Prevalence trends*

Two sources of information are available to estimate recent trends in CVD prevalence in England: the Quality and Outcomes Framework (please refer to footnote in section 1.1.3.3 on page 8 for a short description), and the nationally representative General Lifestyle Survey[4]. Based on the former, CHD prevalence in England has been slowly declining since 2004/05, while the prevalence of stroke has been marginally increasing over the same period.[5] Stroke prevalence trends agree with Lee *et al.* findings, who reported a 12.5 % increase between 1999 and 2008, albeit they used the more selected population of General Practice Research Database.[13]

According to the self-reported prevalence from the General Lifestyle Survey, CVD prevalence increased between 1988 and 2002, then plateaued until 2005 and has been declining since. From the same survey, AMI prevalence has been slowly declining since 1988, while stroke prevalence has remained more or less steady.[5] The two methods to monitor prevalence are fundamentally different and any direct comparisons would be problematic.

---

3 Recently renamed to Clinical Practice Research Datalink.
4 General Lifestyle Survey was named as 'General Household Survey' before 2006.
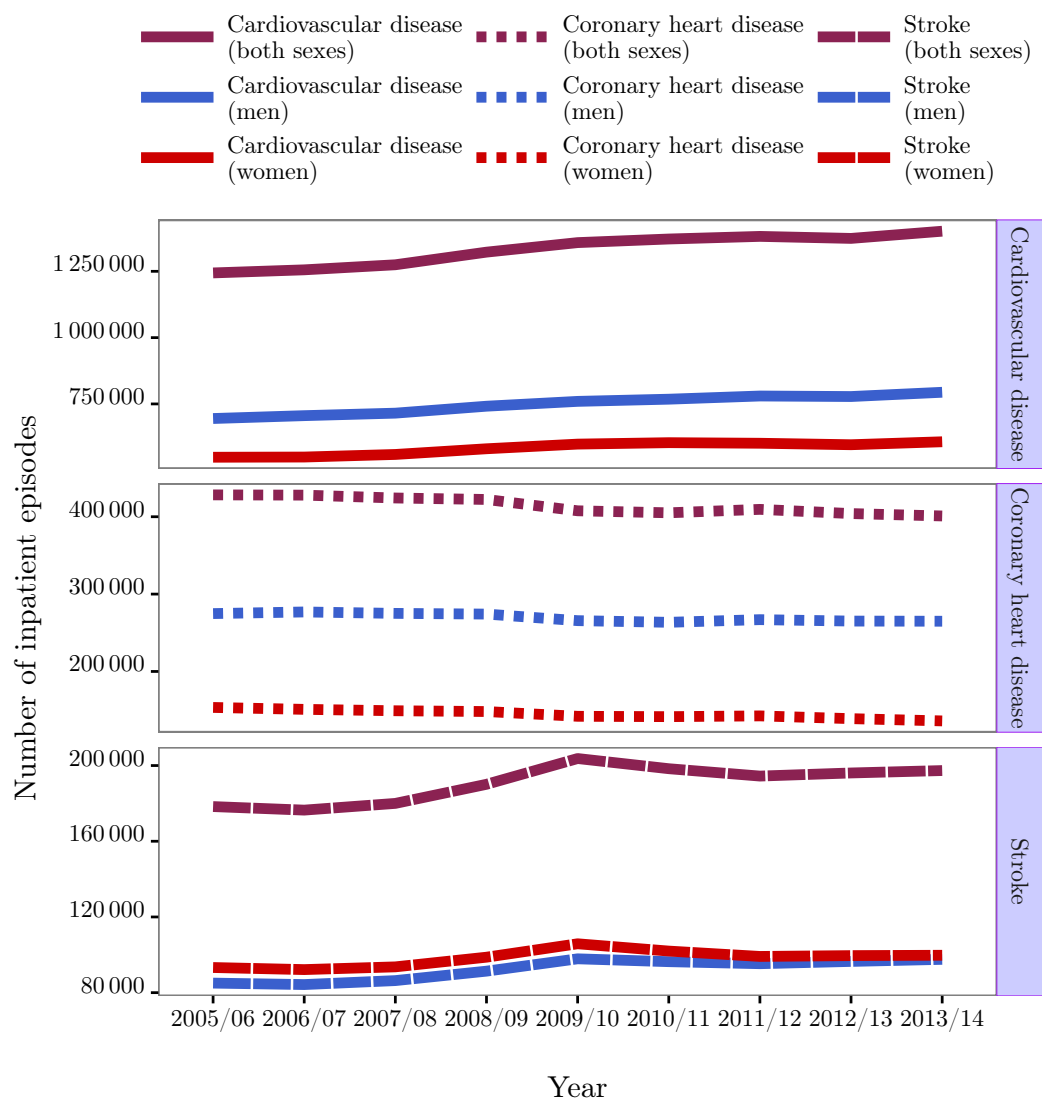
Figure 1.3: Inpatient episodes for cardiovascular disease, coronary heart disease, and stroke in England since 2005/06. Please note that the vertical axis is not starting from 0 and has a different scale for each disease. Data source: Cardiovascular Disease Statistics 2015 report.[5]

Cancer is a heterogeneous group of diseases that can affect any part of the body and is characterised by the creation and growth of invasive mutant cells.[14] Cancer is the second leading cause of death worldwide, responsible for more than 8.2 million deaths in 2013, an increase of 45.6 % since 1990.[2] As with CVD, the ageing of the global population is the main reason behind the increase in cancer related deaths. The age-standardised mortality declined by 14.7 % over the same period. A decrease smaller by 7.3 % compared to the observed decrease in age-standardised CVD mortality (section 1.1.3.1 on page 7).

In England and Wales, cancer is the leading cause of death among both men and women. Cancer accounted for 32 % of all male deaths and 27 % of all female deaths in 2014. The recent age-standardised cancer mortality rate trends have been decreasing for all ages, except for the over 80 age group.[15, 16]

Unlike cancer mortality rate trends, the age-standardised incidence rate trends have been increasing for all age groups since 1979 in GB (figure 1.4 on the next page).[16] The opposite directions of incidence and mortality rate trends directly reflect improvements in the survival of cancer patients over the recent decades. It is worth noting here that unlike CVD incidence, cancer incidence in England (and the rest of the UK) is accurately recorded through cancer registries. The quality of English cancer registration is excellent and there are processes in place to ensure its validity.[17]

In my thesis, I will focus on two specific cancers: 1. lung cancer and 2. gastric cancer. The decision to include only two cancers was pragmatic, based on my limited time and resources. My ambition was for these cancers to be a proof of concept that multiple NCDs with different epidemiologies can coexist in the model. However, the choice of these two specific cancers was far from haphazard; it was based on their significant burden, the fact that they share common determinants with CVD, and their interesting epidemiology. As I will describe in the following paragraphs, lung cancer poses a huge burden on the population and its incidence trend has been increasing for years. On the other hand, gastric cancer incidence is declining almost as fast as CVD.

1.2.1   *Overview of lung cancer epidemiology*

Lung cancer causes more deaths globally than any other cancer.[2, 18] Four major histologic types of lung cancer have been identified so far and this differentiation is important from a clinical perspective. Nevertheless, from an epidemiological perspective all subtypes share common determinants and have a similar prognosis with only nuanced differences. Hence, for the purpose of this thesis I will consider lung cancer as a homogeneous disease.

1.2.1.1   *Mortality and trends*

Lung cancer is the most common cause of cancer deaths in England for both sexes. It was also estimated as the fifth cause of disability-adjusted life years (DALYs) in 2013. In 2014, 15 856 men and 12 993 women died of lung cancer in England. Almost 9 in 10 of
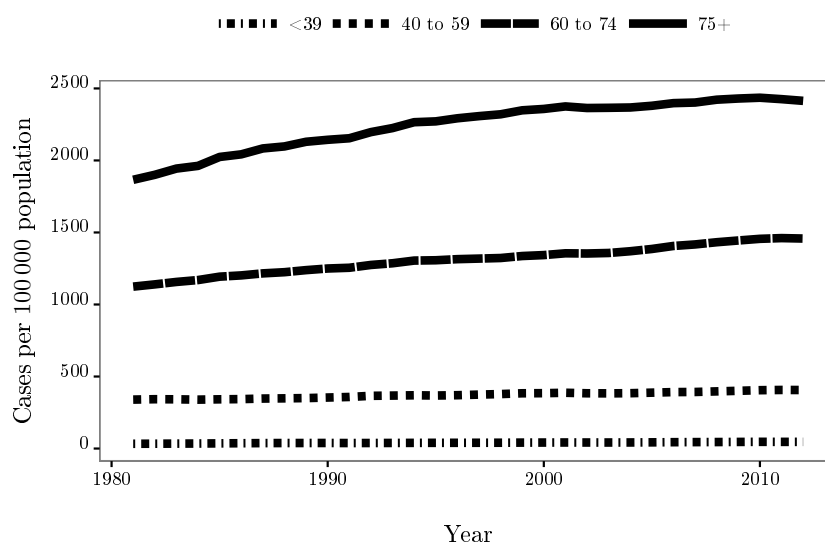
Figure 1.4: All cancers (excluding non-melanoma skin cancer) incidence rates by age group in Great Britain, 1979 - 2012. Directly age-standardised against the European standard population.[10] Data source: Cancer Research United Kingdom.[16]

these deaths occurred in ages older than 60 years. On the contrary, lung cancer deaths are rare in ages under 45. The European age-standardised mortality rate for men was 72.9 per 100 000 population and 48.4 per 100 000 population for women. Similarly to CVD mortality, England has the lowest age-standardised mortality rate among the four UK countries.[19, 20]

Lung cancer mortality time trends in the UK have different patterns by sex, primarily reflecting differences in smoking histories (please refer to section 1.2.1.2). Specifically, the age-standardised mortality rate for men has been declining sharply since the early 1970s. On the contrary, the age-standardised mortality rate for women has been increasing over the same period. The increase was faster between early 1970s and late 1980s and has slowed down since.[19]

1.2.1.2   *Incidence and trends*

Lung cancer incidence patterns are similar to its mortality patterns, with the majority of cases diagnosed in ages over 60 years. Lung cancer is the second most common cancer both for men and women, and the third when both sexes are considered together. In 2013, 19 830 men and 16 823 women were diagnosed with lung cancer, in England; corresponding to 92.5 and 64.4 age-standardised incidence rate per 100 000 population, respectively.[19]

The age-standardised lung cancer incidence rate for both sexes combined decreased by more than 17 % between 1979 and 2003, but a slow upward trend has been observed since. This time trend is the net effect of two opposite direction time trends; the age-standardised incidence rate for men has been decreasing since 1979, while it has been constantly increas-

ing for women over the same period. These differences in time trends reflect the different smoking histories of men and women in the UK (please refer to section 1.3.1 on page 18). In particular, smoking prevalence has been steadily dropping at least since 1948 for men, while it had been increasing until 1970 for women before declining since. The period of smoking expansion before 1970 for women is responsible for the observed increasing trends in lung cancer incidence more than four decades later.[19]

1.2.1.3   *Survival and trends*

Unfortunately, lung cancer has a grave prognosis. The similarity in incidence and mortality patterns is actually the result of poor survival. In 2010, 5-year survival from lung cancer was no more than 10 %. Survival tends to be slightly better among women, and for younger patients.[19]

Survival from lung cancer has shown little improvement over the last four decades. 5-year survival has only improved by about 4 % (absolute) for men and 7 % (absolute) for women since 1970. One year survival improvement was about three times higher over the same period. Hence, treatment advancements over the recent decades may have extended survival, but did not substantially improve the cure rate.[19, 21]

1.2.1.4   *Prevalence and trends*

Information regarding lung cancer prevalence is scarce and occasionally contradictory, as with many other cancers, mainly because there is no clear 'case definition'. For example, Maddams *et al.* used cancer registry data and estimated that 63 522 people were living with lung cancer in the UK, in 2008.[22] On the other hand, the British Lung Foundation used data from the Health Improvement Network database[5] records and estimated that approximately 81 800 people diagnosed with lung cancer were living in the UK in 2008. The same institution estimated that the lung cancer prevalence is rising over time.[23]

1.2.2   *Overview of gastric cancer epidemiology*

Gastric cancer is the fifth most common cancer and the third most common cancer cause of death worldwide.[18] Its age-standardised mortality rate dropped by more than 36 % (relative) between 1990 and 2013.[2] As with lung cancer, gastric cancer has several taxonomies that are important in clinical practice. One specific taxonomy is also important from an epidemiological perspective. Tumours that are located closer to the oesophagus are known as gastric cardia cancer and appear to have important differences from the non-cardia gastric cancer. The former type appears to be more common in high-income countries (including the UK) and its age-standardised incidence rate is increasing worldwide. In contrast, non-cardia gastric cancer is the most common subtype worldwide and its age-standardised incidence rate is declining.[24]

---

5 The database contains more than 12 million patient records from 591 General Practitioner surgeries.
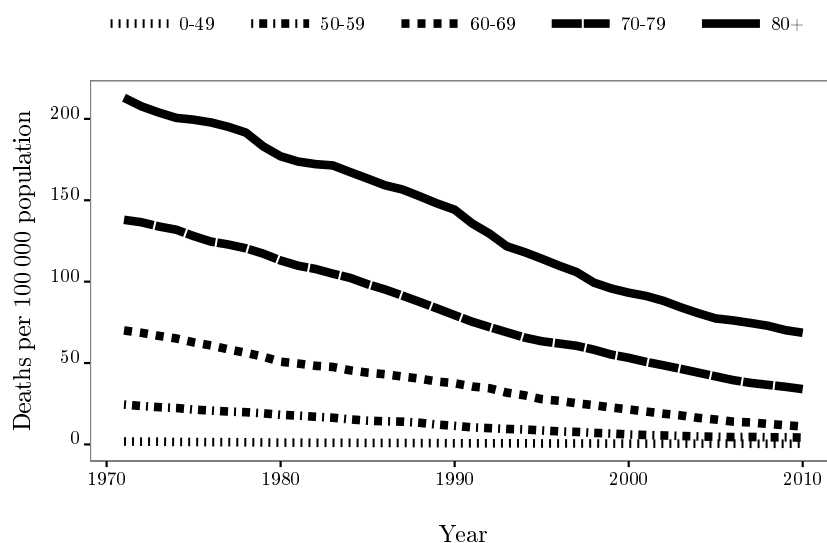
Figure 1.5: Gastric cancer mortality rate trends in the United Kingdom, 1971 - 2010. Directly age-standardised against the European standard population.[10] Data source: Cancer Research United Kingdom.[25]

1.2.2.1 *Mortality and trends*

In 2014, gastric cancer was the eighth most common cause of cancer death for men and the 13th for women in the UK. In England, 2330 men and 1298 women died of gastric cancer over the same year. These deaths correspond to a standardised mortality rate of 10.9 per 100 000 for men and 4.7 per 100 000 for women. Once again, this is the lowest among the four UK countries. Finally, gastric cancer mortality increases exponentially with age.[25]

The age-standardised mortality rate has been decreasing since 1970 in the UK and the overall reduction was around 77 % (relative) faster than the global trends. The decrease has been observed in all age groups and it was faster for those aged 60 to 69 and slower for those aged over 80 (figure 1.5).[25]

1.2.2.2 *Incidence and trends*

As with lung cancer, incidence of gastric cancer shows similar patterns to mortality. In 2013, 3674 men and 1967 women were diagnosed with gastric cancer in England, corresponding to 17.3 per 100 000 and 7.3 per 100 000 age-standardised incidence ratio, respectively. The majority of these tumours occurred in cardia and were diagnosed in people aged over 75.[25]

The age-standardised incidence rate of gastric cancer has been declining since the late 1970s by about 62 % (relative) in GB. The decline was similar for both sexes and explains at large the observed decline in mortality. The fastest decline has been observed in the 50 to 69 age group and the slowest in the 25 to 49 age group.[25] Unlike the global trends, the incidence decline has been observed for both cardia and non-cardia gastric cancer.[26]

1.2.2.3 *Survival and trends*

Gastric cancer prognosis is marginally better than that for lung cancer. Age-standardised 5-year survival from gastric cancer is less than 20 %. Overall, survival is better for younger ages and for women.[25]

Survival has been improved substantially since 1970. Age-standardised 5-year survival was around 5 % in the early 1970s and was almost four times higher in 2011. Similar relative improvement has been observed in one year survival.[21, 25]

1.2.2.4 *Prevalence*

The difficulties in accurate estimation of gastric cancer are similar to lung cancer. The now outdated EUROPREVAL study suggested that approximately 45 per 100 000 men and 26 per 100 000 women were living with a diagnosis of gastric cancer in 1992 in England.[27] More recent estimates, from the National Cancer Intelligence Network, calculated that in 2006 7048 men and 3745 women were living in England with a diagnosis of gastric cancer within ten years of diagnosis. These correspond to a crude estimate of 28.3 for men and 14.5 per 100 000 for women and an age-standardised estimate of 23.0 and 9.6 per 100 000, respectively.[28]

1.3 AN OVERVIEW OF NON-COMMUNICABLE DISEASE-RELATED MODIFIABLE RISK FACTORS EPIDEMIOLOGY

It is apparent from the previous paragraphs that the incidence of the most important NCDs has changed substantially over time, driving mortality trends. Undoubtedly, the ageing of the population has been a powerful driver of crude incidence trends. However, substantive changes have been observed to the age-standardised incidence rates, as well. The shaping force behind the evolving age-standardised NCDs incidence patterns is trends in the associated risk factors. It is well accepted now that most NCDs share four common behavioural risk factors: tobacco, unhealthy diet, physical inactivity, and alcohol excess.[20, 29–33] In England, these factors accounted for about 25 % of the total DALYs in 2013; their contribution was even higher in CVD DALYs (55 %) and cancer DALYs (34 %). Unhealthy diet, and tobacco specifically were the leading contributors to DALYs overall.[20] Therefore, patterns and trends in population exposure to these risk factors will powerfully influence NCDs incidence and mortality.

In the following paragraphs, I will describe the patterns and trends in modifiable risk factors in England. Moreover, I will briefly summarise the available evidence, regarding their associations with each of the diseases that have been presented in this thesis. I will present only modifiable risk factors, because by definition non-modifiable ones are irrelevant to prevention, and I will only briefly mention for completeness modifiable risk factors that were not examined in this thesis. Whenever relevant reports, systematic reviews, or meta-analyses exist to summarise the evidence, I will refer to them instead of the primary studies.

1.3.1   *Tobacco and smoking*

Evidence that tobacco and primarily tobacco smoking are associated with unfavourable health outcomes has been apparent since the 1930s.[34–36] In fact, many modern epidemiological concepts and methods were initially developed to establish the causal relation between tobacco and diseases. Numerous high quality epidemiological and laboratory studies have confirmed the causal link between smoking, CVD, lung, and gastric cancer. The link has also been confirmed for environmental tobacco smoking, CVD, and lung cancer. A small proportion of the excess risk of smoking for CVD is mediated through an adverse effect on blood pressure.[37] The evidence has been summarised in reports from the World Health Organisation (WHO), the United States (US) Surgeon General, and the International Agency for Research on Cancer.[38–40] A recent study has expanded the list of diseases linked to smoking even further.[41]

1.3.1.1   *Risk reversibility*

Risk reversibility is important for primary prevention policies. It refers to whether and how fast the excess risk declines or reverts back to zero after the exposure ceases. Several studies have shown that the risk gradually decreases for ex-smokers, although the rate of decline varies by disease and study. For lung cancer, the risk drops fast the first years after smoking cessation and then the rate of decrease slows down before it levels off about 30 years since cessation. A small residual risk remains even after 35 years post cessation. For CVD, the pattern is similar but shorter; the excess risk disappears about 15 years post cessation with no residue.[38] Interestingly, for stroke specifically the risk completely eclipses five years post cessation in the Framingham study.[42]

1.3.1.2   *Patterns and trends*

In 2014, about 19 % of the population aged over 16 were active smokers in GB. A similar percentage of secondary school pupils reported that they had tried smoking at least once. Smoking was moderately more prevalent among men than women, 20 % versus 17 %. It was also more prevalent in younger ages. Smoking peaked at the age group of 25 to 34 and gradually declined in older age groups; smoking prevalence was around 11 % for those aged 60 and over. In 2013, about 31 % of men and 26 % of women reported exposure to environmental tobacco smoking. For secondary school pupils, the prevalence of environmental tobacco smoking rose to 64 %. Smoking prevalence has dropped by about 60 % since the 1970s. Nevertheless, smoking patterns showed little change over the last decade.[43] It is worth noting that estimates for England from a different source, the HSE, are almost identical.[44]

1.3.2   *Unhealthy diet*

Unhealthy diet is a generic, multidimensional term, that amalgamates evidence and controversies on a spectrum of diet related issues. Nutrition science has developed rapidly

in the past few decades; however controversies remain. In a provocative study, Schoenfeld *et al.* randomly selected 50 common ingredients from a cookbook and reported that 40 were apparently associated with increased cancer risk in peer-reviewed studies. Unsurprisingly, most of these associations disappeared in subsequent meta-analyses.[45] Despite the controversies, there is now convincing evidence to suggest that a healthy diet is rich in fruits and non-starchy vegetables, nuts, legumes, dietary fibre, and minimally processed foods and does not contain processed meat, trans fats, excess salt, and excess sugars.[31, 46, 47] In this thesis, I will focus on two aspects of unhealthy diet: low fruit and vegetable consumption;[6] and excess salt consumption.

1.3.3 *Low fruit and vegetable consumption*

The evidence regarding the causal links of inadequate fruit and vegetable consumption with cardiovascular disease and cancers is summarised in the reports from the World Cancer Research Fund and the WHO.[31, 48] However, uncertainty still exists regarding which specific fruit or vegetable may have the most protective effect and whether there is a ceiling effect after which no further benefits occur. Generally, most national dietary guidelines recommend at least 400 g of mixed fruit and vegetables daily and mention no maximum limit. For CVD a recent meta-analysis from Wang *et al.* showed that a ceiling effect may exist after 400 g of daily intake.[49] However, this meta-analysis did not include a large and well designed study in England that suggested significant benefits even for higher consumptions.[50, 51] For lung cancer, two similar meta-analyses concluded that a ceiling effect may exist after 400 g of daily intake.[52, 53] Finally, for gastric cancer a recent meta-analysis for the Continuous Update Project found limited but suggestive evidence that up to 150 g of fruit intake has a protective effect, which levels off for higher consumption.[24]

1.3.3.1 *Risk reversibility*

The evidence is limited regarding risk reversibility. For CVD, one randomised control trial for CHD secondary prevention showed that risk can decrease within a year of an increase in fruit and vegetable consumption and reduced fat consumption. Natural experiments in East Germany, Hungary, Romania, and Poland during the socioeconomic transformation in the 1990s provide some evidence that risk decline can be observed within two to four years. With regard to lung cancer, evidence is even more limited. Two observational studies are suggestive of a longer latent period of four to eight years before a decrease in incidence can be observed, after an increase in fruit and vegetable intake.[48]

1.3.3.2 *Patterns and trends*

Findings from the nationally representative National Diet and Nutrition Survey (NDNS) revealed that in the period 2008 to 2012, only about 30 % of the adult population in the UK was consuming more than 400 g of fruit and vegetables daily.[7] The mean daily con-

---

6 For convenience, from now on I will use the term 'vegetable' to refer to non-starchy vegetable.
7 400 g of fruit and vegetables is currently the national target, also known as the '5 a day'.

sumption for adults was about 350 g and consumption increased by age. However, NDNS participants under-reported overall calorie intake indicating possible social desirability bias.[54] For England, the HSE2013 produced lower estimates. Specifically, the mean fruit and vegetable consumption was found to be around 290 g and only about 28 % of women and 25 % of men had an intake of more than 400 g. HSE is contacted on an annual basis with similar methodology and trends can be safely extracted. It appears that fruit and vegetable consumption has been more or less stable since 2001 with the possible exception of a small peak in 2006.[55]

1.3.4   *Excess salt consumption*

Excess dietary salt consumption has been linked to an increased risk for CVD and gastric cancer.[24, 56, 57] For CVD, the excess risk appears to be mainly mediated through the deleterious effect of excess salt consumption on blood pressure.[58, 59] The pathophysiological mechanisms that link excess salt consumption with the increased risk for gastric cancer are less clear. Some experimental studies showed increased inflammation of gastric mucosa, caused by high intragastric sodium concentrations, that leads to increased cell mutations. Other researchers suggest that a high salt diet facilitates gastric colonisation by Helicobacter pylori, a widely accepted risk factor for gastric cancer, through changes in the viscosity of the gastric mucous barrier.[57]

There is some controversy regarding the optimal level of salt consumption.[60] The WHO and the UK national guidelines recommend a daily salt intake of less than 5 g and 6 g, respectively.[61, 62] Some researchers claim that salt consumption lower than 7.5 g can actually increase the risk of CVD and overall mortality.[63, 64] However, it appears that this argument is based on biased measurement methodology.[65] A recent discussion on the subject can be found in Mozaffarian *et al.* who concluded that the optimal level of salt consumption below which no health gains have been observed is somewhere in the range of 1.5 g/d to 5.9 g/d.[59]

1.3.4.1   *Risk reversibility*

Evidence that directly links salt risk reversibility to CVD mortality or morbidity outcomes is lacking. A meta-analysis of several randomised control trials that tested low salt diets was underpowered and therefore inconclusive.[66] In comparison, a plethora exists on the effect of low salt diet on systolic blood pressure (SBP), which appears to happen within weeks.[58, 59, 67] Finally, to my knowledge there is no convincing evidence regarding risk reversibility for gastric cancer.

1.3.4.2   *Patterns and trends*

In the UK, about 70 % of dietary salt consumption comes from processed food.[62] In 2011, the NDNS showed that the mean daily salt consumption among adults aged 19 to 64 years was 8.1 g/d. A significant sex difference was observed with men having a mean estimated intake of 9.3 g/d compared to 6.8 g/d for women. Furthermore, salt consumption appeared

to decrease with age. Overall, less than 30 % of the population achieved the national target of less than 6 g/d. Between 2001 and 2011 the mean salt consumption dropped from 9.5 g/d to 8.1 g/d. The reduction was observed for both sexes and all age groups.[68]

1.3.5 *Physical inactivity*

Despite the practical difficulties to accurately measure the level of activity of an individual, a substantial body of evidence suggests the association of physical inactivity with an increased risk of CVD, and all cancer morbidity and mortality. [31, 69, 70] However, there is limited and inconclusive evidence to suggest an association of physical inactivity with an increased risk specifically for gastric cancer; and limited but suggestive for lung cancer.[24, 31, 69] Several plausible biological mechanisms exist to explain how physical activity reduces cardiovascular risk[8] through beneficial effects on blood pressure, glucose, and lipid profile among others.[69]

1.3.5.1 *Risk reversibility*

Some evidence exists from randomised control trials to suggest that risk reversibility occurs within a couple of years from an increase in physical activity, at least for CHD. Evidence about risk reversibility for stroke is very limited.[69]

1.3.5.2 *Patterns and trends in England*

UK guidelines recommend at least 150 minutes of moderate intensity activity in bouts of 10 minutes or more, weekly for adults.[71] Findings from HSE2012 suggest that about 67 % of men and 55 % of women have reported that they meet the national targets. Younger participants with normal weight were more likely to report that they were more active. Although physical activity national guidelines and the physical activity questionnaire of the HSE have been updated several times since 2001, it seems that the mean activity level in England has been moderately increased.[72]

1.3.6 *High body mass index*

Body mass index (BMI) is a measure of body adiposity that is calculated by the body weight (in kg) divided by the squared body height (in m²). Strong evidence exists that BMI higher than 20 kg/m² is an independent and dose dependent risk for CVD at the population level.[9] A large part of the excess risk from high BMI is mediated through adverse effects on blood pressure, lipid profile, and glycaemic control.[73, 74] High BMI has been also associated with an increased risk for certain cancers; however, until recently lung and gastric cancers were not among them.[31] In 2016, the Continuous Update Project of the World Cancer Research Fund International ruled that enough evidence exists to suggest that BMI is a

---

8 Although, some uncertainty exist for haemorrhagic stroke.
9 For completeness, BMI lower than about 17.5 kg/m² is also associated with increased overall mortality.

risk factor for gastric cardia cancer specifically.[24] Surprisingly, some recent evidence suggests that high BMI may decrease the risk for lung cancer.[75, 76]

1.3.6.1   *Risk reversibility*

Enough evidence exists to suggest that risk reversibility for CVD after reductions in BMI is happening well within five years. In fact, favourable changes in blood pressure, lipid profile, and glycaemic control have been observed within days or weeks.[73] Evidence on risk reversibility for gastric cancer is non-existent; however, some evidence from post-menopausal breast cancer trials suggests that premenopausal weight reduction may affect risk of post-menopausal breast cancer. Therefore, the lag time for gastric cardia cancer may well be longer than the lag time for CVD.[24, 73]

1.3.6.2   *Patterns and trends in England*

Findings from the HSE suggest that in 2014 the mean BMI was $27.2\,\text{kg/m}^2$ for adult men and women in England. BMI peaks in the 55 to 64 age group for both sexes and then gradually declines for older age groups. Since 1993, mean BMI has increased from $25.8\,\text{kg/m}^2$. The increase has been observed for all age groups, however it has substantially slowed down since 2006.[77]

1.3.7   *High blood pressure*

SBP higher than 115 mmHg or diastolic blood pressure higher than 75 mmHg has been linked with an increased risk for CVD. Plausible biological mechanisms have been identified that can explain the adverse effects of high blood pressure through the stiffening of the arteries. No causal relation with any type of cancer has been observed so far.[78, 79]

1.3.7.1   *Risk reversibility*

Extensive evidence exists from randomised clinical trials that have been conducted with blood pressure related interventions. For CHD about two thirds of the risk are reversible within three to five years. It appears that some residual risk remains even after five years. In contrast, for stroke the risk is fully reversible within three to five years from blood pressure reduction.[79]

1.3.7.2   *Patterns and trends in England*

As I have already briefly stated in previous paragraphs, smoking, salt, high BMI, physical inactivity, and possibly fruit and vegetable consumption mediate their risk through blood pressure. Hence, trends of these behavioural risk factors affect blood pressure trends. HSE provides useful information about the exposure of the population to hypertension[10]. According to HSE findings, in 2014 the prevalence of hypertension was 32.4 % for adult men

---

10 HSE defines hypertension as having blood pressure higher than 140/90 mmHg or being on antihypertensive medication.

and 26.9 % for adult women and sharply increased with age for both sexes. Hypertension prevalence among adult participants has dropped by 1 % (absolute) since 2003, and by 2.1 % (absolute) between 1998 and 2003.[44, 80]

1.3.8   *Total serum cholesterol*

Total serum cholesterol is an independent risk factor for CVD, although the association with haemorrhagic stroke is uncertain. A large number of cohort studies showed that participants with serum cholesterol higher than 3.8 mmol/l have a higher risk for CVD. Elevated serum cholesterol is a well established and substantial risk factor for atherosclerosis.[81, 82]

1.3.8.1   *Risk reversibility*

As with SBP, several randomised control trials have been conducted with serum cholesterol interventions. From their results, it appears that risk reduction for CHD is observable within two years from cholesterol reduction, and essentially all the excess risk is reversed within five years. Evidence for stroke is less clear.[82]

1.3.8.2   *Patterns and trends in England*

In 2011, mean total cholesterol was 5.1 mmol/l among adult men and 5.2 mmol/l among adult women in England. Mean cholesterol peaks in the age group 45 to 54 for men and 55 to 64 for women and declines again in older ages. Over the last two decades, mean cholesterol remained stable between 1998 and 2003, and then declined by about 0.5 mmol/l between 2003 and 2011.[83]However, evidence suggests that mean total cholesterol in the population had been declining for at least two decades before that.[84]

In summary, three key concepts have emerged so far: 1. the observed incidence trends of CVD, lung, and gastric cancer emanate from the interplay of the trends in the aforementioned modifiable risk factors; 2. past exposures influence current disease incidence and current exposures will influence future disease incidence, because of the lag times between exposure and disease; 3. the lag time for CVD appears to be shorter than five years, while the lag time for cancers appears to be longer, albeit evidence is limited.

1.4   SOCIOECONOMIC HEALTH INEQUALITIES AND THE SOCIAL DETERMINANTS OF HEALTH

Kawachi *et al.* defined health inequality as "…the generic term used to designate differences, variations, and disparities in the health achievements of individuals and groups".[85] In England, health inequalities remain and are projected to increase in the future. A large proportion of health inequalities can be explained by socioeconomic inequalities and it is evident that poor health disproportionately burdens the more disadvantaged in our society.[20, 86] The term 'health inequities' is a politically charged term that emphasises the

unnecessary and unjust nature of some of the observed health inequalities. Some believe that all socioeconomic inequalities in health are unjust and preventable;[87] although this is not universally accepted and the debate is summarised in Kawachi *et al.*[85] In this thesis, I will use the terms 'inequity' when I refer to resources, and 'inequality' when I refer to the outcomes.

Socioeconomic inequalities in health can be explained, at least partly, by socioeconomic gradients in the prevalence of modifiable risk factors that have been observed in England. Specifically, smoking, unhealthy diet, BMI, physical inactivity, and hypertension increase their prevalence by deprivation.[11, 43, 88–91] Interestingly, no socioeconomic gradient has been observed recently for serum total cholesterol.[83] Consequently, the socioeconomic gradients in risk factor exposures are translated to gradients in the burdens of CVD, lung, and gastric cancer. Although CVD incidence is largely unknown, premature CVD mortality shows a substantial socioeconomic gradient with those in a routine occupation to have about three times higher CVD mortality rate than those in a managerial or professional occupation (figure 1.6 on the next page).[92] Among cancers, lung cancer is the one with the most unequal burden both for incidence and mortality.

In the period 2006 - 2010, 11 700 cases and 9900 deaths from lung cancer could have been prevented every year, if all socioeconomic quintiles could experience the same lung cancer incidence and mortality rates as the least deprived fifth in England. Gastric cancer showed similar patterns and had the third highest excess cases and deaths over the same period; 1400 and 1000, respectively.[93] Despite the recent improvements in the incidence and mortality of these diseases, the socioeconomic gradients persist and may have even increased, with the notable exception of gastric cancer in men.[92, 93]

1.4.1   *Social determinants of health*

The socioeconomic gradients in CVD and certain cancers may be somewhat explained by the socioeconomic gradients in risk factor exposures. Yet the question remains about what causes the latter; what Marmot terms "the causes of the causes". The social determinants of health approach offers an explanation regarding the unequal distribution of risk exposures and disease burden in different socioeconomic groups. According to this approach, socioeconomic inequalities in health are generated by socioeconomic inequalities in the population. The social determinants of health approach has been adopted by the WHO and the recent Marmot review of health inequalities in the UK.[87, 97, 98]

The social determinants of health are exceptionally summarised in a rainbow graph by Dahlgren *et al.*, which illustrates all major determinants of health in a hierarchical manner (figure 1.7 on page 26). The main notion of the graph is that every layer influences its inner layers and gets influenced by its outer layers. Hence, in the centre circle there is the individuals with their individual biological traits. Many of the traits are non-modifiable, like age or genetic make up. Nevertheless, some can be influenced by the outer layers, like SBP or total cholesterol. The next layer consists of the behaviours that the individuals choose to adopt. For instance, smoking or unhealthy diet. These behaviours influence the centre circle and are influenced by the outer layers. Individual choices are made within a
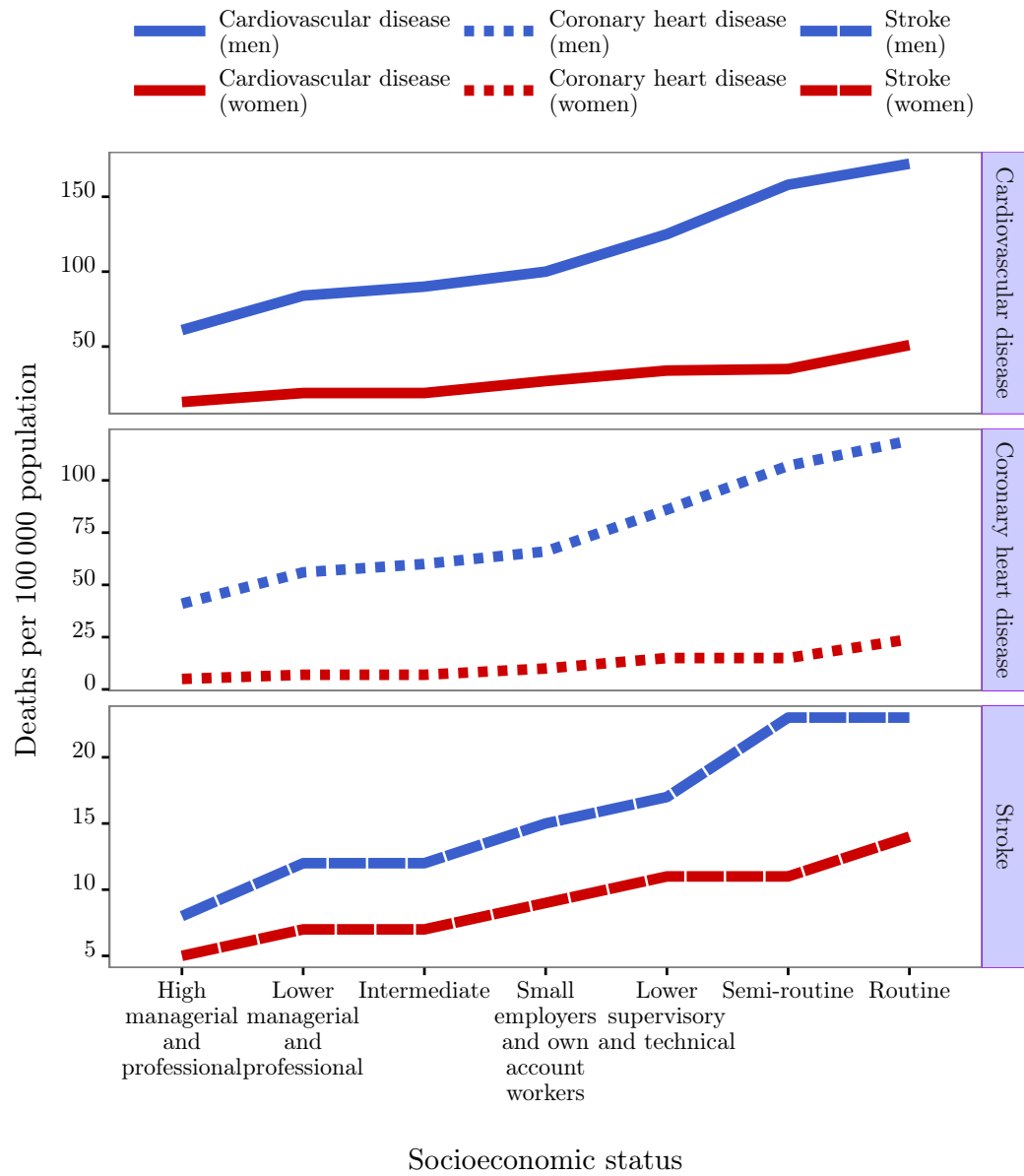
Figure 1.6: Cardiovascular disease mortality rates by socioeconomic status and sex, 2001 - 2003 in England and Wales. Directly age-standardised to the European Standard Population.[10] Socioeconomic status categories were defined by the National Statistics Socio Economic Classification.[94] Age range for men is 25 to 64. Age range for women is 25 to 59. Data sources: Office for National Statistics mortality register.[95, 96]
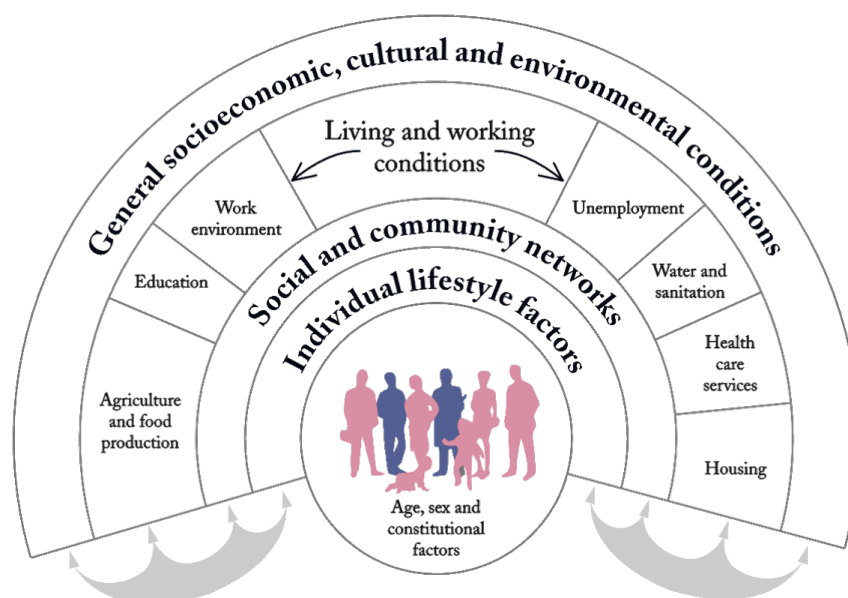
Figure 1.7: The social determinants of health rainbow. Adapted from Dahlgren *et al.*[99]

context of an interplay between socioeconomic, cultural, and environmental conditions at macro, meso, and micro level[11] throughout the life course.[87, 99]

1.4.2   *The social production of disease model*

The social determinants of health approach provides a useful conceptual framework spanning much of the force of morbidity and mortality in society. However, the question remains regarding through which specific pathways and mechanisms socioeconomic inequalities generate health inequalities. Many theories have been proposed on the subject over the years. The Diderichsen theoretical model of the social production of disease has captured key elements of these theories (figure 1.8 on the next page).[100] It offers four concurrent pathways that link the broader socioeconomic context to individual level health: 1. through social stratification itself; 2. differential exposure to risk factors due to social stratification; 3. differential vulnerability at the same level of exposure, and; 4. differential consequences of ill health. Crucially, it also contains a positive feedback loop that allows health inequalities to feed back to socioeconomic inequalities.

PATHWAY 1: SOCIAL STRATIFICATION   This is the cornerstone pathway in the model and links it to the social determinants of health approach. The unequal social structures lead to unequal social position of individuals. However, it is perhaps unrealistic to expect that this might improve in the foreseeable future.

---

11 Notably income, power, housing and education.

SOCIETY   INDIVIDUAL
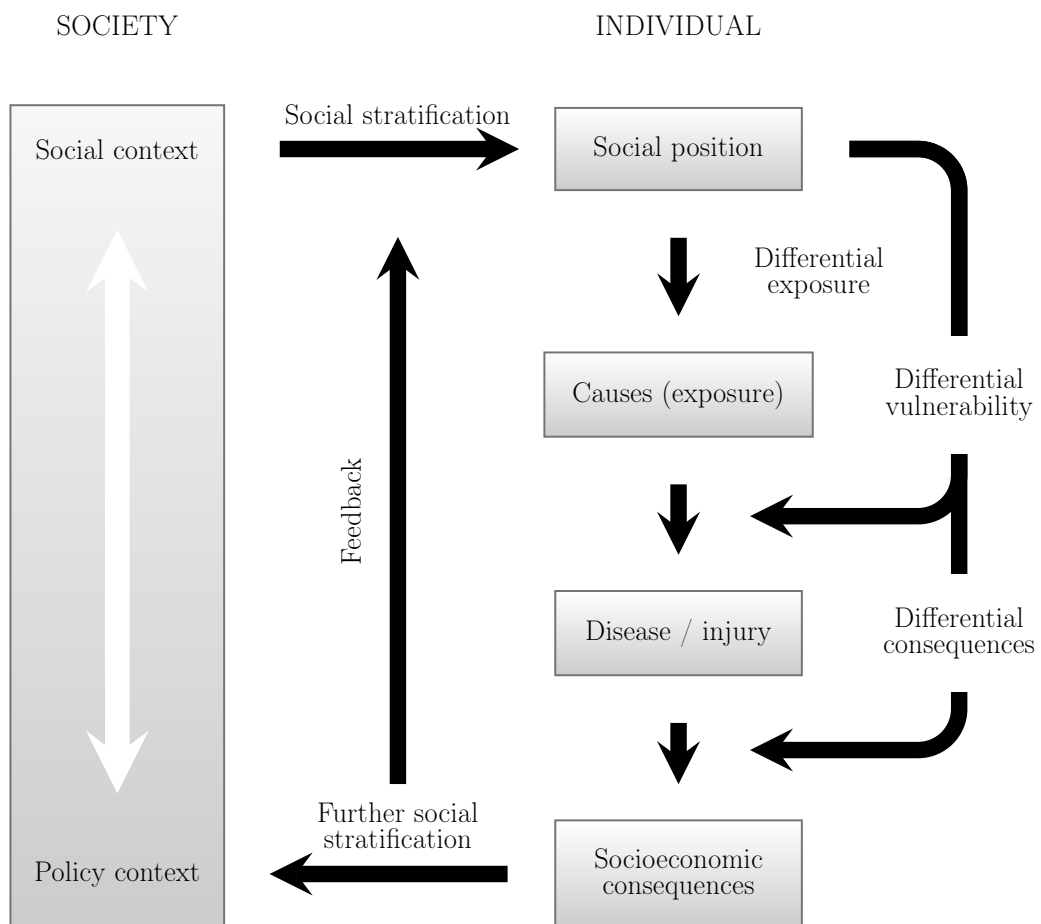
Figure 1.8: Diderichsen model of social production of disease. Adapted from Diderichsen *et al.*[100]

PATHWAY 2: DIFFERENTIAL EXPOSURE   Individuals in different social positions experience different exposure to risk factors. Risks factors may be behavioural like smoking, or environmental like air pollution or occupational hazards. From a public health perspective this is the most important pathway, because it also affects the magnitude of the remaining two pathways. In one study, Lynch *et al.* studied a cohort of 2272 Finnish men and found that the differential exposure to behavioural and biological risk factors could explain a substantial proportion of the observed socioeconomic gradient in CVD mortality.[101]

PATHWAY 3: DIFFERENTIAL VULNERABILITY   This is perhaps the most subtle of the four pathways. It hypothesises that for the same level of exposure to a risk factor, individuals in different socioeconomic positions experience different health effects. This may occur through effect modification of exposures. This is a phenomenon in which the effect of an exposure is amplified or attenuated due to the differential exposure level of another risk factor. For example, the risk of CVD due to hypertension may be amplified if the individual is also a smoker; because smoking is more prevalent in lower socioeconomic groups that increases the vulnerability of these groups to hypertension. Effect modification has been shown to apply to major CVD risk factors and their effects appears to be multiplicative.[102, 103] Hence, the clustering of multiple risk factors in the lower socioeconomic groups multiply their effects.

PATHWAY 4: DIFFERENTIAL CONSEQUENCES   Unlike the previous two pathways that are related to primary prevention, this pathway applies after the development of the disease. Disease consequences are also influenced by socioeconomic position. Differential consequences have been clearly demonstrated in CVD survival,[104–106] and cancer outcomes and survival.[107–110] Furthermore, this pathway is essential for the feedback loop. Individuals and families in lower socioeconomic positions have fewer resources to counteract the consequences of ill health, which may lead to unemployment, reduced income, and less years in education; therefore, ill health may worsen their socioeconomic position even further.

One of the strengths of the Diderichsen model is that it acknowledges that policy plays an important role in the social production of disease. First, it recognises that policy and social context interact. Second, it offers a practical framework to policy makers and planners for policy design. Policies that aim to negate or weaken any of the four pathways can potentially tackle socioeconomic inequalities in health. The more upstream the target pathway is, i. e. social stratification or differential exposure pathways, the more equitable a policy can be. In contrast, policies that strengthen these pathways generate further health inequalities. The Diderichsen model is now widely accepted and has been used by the WHO Commission for Social Determinants of Health.[97] Its conceptual clarity makes it also ideal to guide data analysis and simulation modelling. I will describe later in the methods chapter how I used this framework to structure my model.

1.4.3 *Preventive interventions and socioeconomic health inequalities*

Diderichsen's model highlights the strong ties of policy with socioeconomic health inequalities. Yet, empirical evidence regarding the impact of preventive policies on inequalities is sparse. A recent systematic review on interventions to promote healthy eating identified 199 relevant studies; however, only 36 of them reported their results stratified by socioeconomic position.[111, 112] Consequently, public health policy makers and planners often have to rely on fragmented or sparse information for their decisions.

1.4.4 *Measures of socioeconomic health inequalities*

Measuring socioeconomic inequalities in health has been a field of research since the 1990s. As an evolving field, there have been numerous debates regarding which is the most appropriate and less biased way of measuring them. One of the first debates was on identifying the most suitable measures of socioeconomic position; occupation, income, education, some combination of these or other measures. It became more complicated when area level deprivation measures became available in some countries, including England. The second controversy was whether a measure of social position is necessary to measure health inequality. Some statistical measures examine the inequality in the distribution of a health measure across the whole population. These measures do not examine whether inequality is preventable, or which groups in the population are burdened disproportionally. These types of measures are particularly popular among health economists. On the contrary, epidemiologists prefer to measure the differences in a health measure across different socioeconomic groups. These debates have been greatly summarised by Regidor and the different measures of socioeconomic inequalities in health have been summarised by Mackenbach *et al.*[113–115]

Finally, the most recent debate is on whether an absolute or relative measure of inequality is most appropriate to be used. This is very similar to the debate about whether absolute or relative risk reduction should be reported in randomised control trials. Harper *et al.* proposed that relative measures should be used when inequalities are the only concern and absolute measures should be used when health inequalities are one of the concerns but other metrics, like population overall health, are also important.[116] Soon after this, Asada used hypothetical examples to show that Harper's *et al.* 'rule of thumb' was oversimplified and could not be applied in many real world examples.[117] It appears that a consensus exists to present both relative and absolute measures of inequality as they highlight different aspects of the same phenomenon.[118]

In conclusion, it is unlikely that everyone would agree in one simple measure for health inequality. However, leaving academic debate aside, practical issues of data availability, the need for comparability and compatibility with other information sources, and the need to communicate results to policy makers that are usually not experts in health inequalities, guide the choice of health inequality measures.

1.4.5   *Tackling health inequities in the United Kingdom*

The UK has a long history of policies targeting health inequalities. Edwin Chadwick in the middle of the nineteenth century undertook an independent inquiry on sanitation after an invitation by the government. The result was 'The sanitary conditions of the labouring population of Great Britain' report and a few years later (and a new government) the 1848 Public Health Act. The Act addressed some of Chadwick's concerns about the living conditions of the labour class in GB. 100 years later the NHS was established to provide universal health care free at the point of service and it was thought that it would tackle health inequalities.[119] The 'Black report' in 1980 came as a wake up call that despite the NHS, socioeconomic inequalities in health persisted and may even have been widened. The report had limited political and policy impact, however it made an impact among researchers and academics.[119, 120] It took another eight years and a new government until the 'Acheson Report' was published that put socioeconomic health inequalities in the political agenda of every government since then.[121]

Today, health inequalities are the focus of many governmental strategies. The Department of Health has set specific objectives and an action plan to reduce health inequalities by 2016.[122] The national strategy for cancer contains specific targets to reduce socioeconomic cancer inequalities.[123] Similarly, the cardiovascular disease outcome strategy also aims to reduce inequalities in CVD.[124] Moreover, the National Institute for Health and Care Excellence (NICE) guidance for Local Authorities and public health practitioners includes recommendations about reducing health inequalities.[125, 126] Furthermore, Marmot *et al.* have stressed that reducing inequities produces both economic and ethical benefits.[127]

Hence, leaving aside academic and practical discussions regarding the unfair or preventable nature of some of the observed health inequalities, it seems the necessity to tackle socioeconomic inequalities in health is now widely accepted and is embodied in state policy documents. Despite the often sparse empirical evidence regarding which specific policies may achieve these targets, public health policies that target smoking, unhealthy diet, and physical inactivity can potentially tackle socioeconomic health inequalities. In the following section I will map the policy options to reduce the burden of NCDs and their likely effect on socioeconomic inequalities.

1.5   PRIMARY PREVENTION TYPOLOGIES

Several attempts have been made so far to categorise primary prevention policies based on some of their characteristics. In general, the benefit of a categorisation framework is that some attributes of a policy can be assumed by analogy even when empirical evidence is lacking. Therefore, for the purpose of this thesis I will describe three different classification frameworks for primary prevention policies based on their potential effectiveness and equity.

1.5.1 *Population-wide versus high-risk prevention*

In the early 1980s Geoffrey Rose revolutionised thinking around primary prevention. His innovative argument was that a moderate decrease in a risk factor exposure across the whole population can be more effective than a larger decrease among those individuals most exposed to the risk (high-risk individuals). For instance, a preventive policy that reduces SBP by 0.3 mmHg across the whole population may prevent more CVD cases than a policy that reduces SBP by 3 mmHg only for high-risk individuals with SBP higher than 160 mmHg.[128–130] Rose's taxonomy was initially described as dichotomous. In real world applications though, it usually represents a continuum. For example a price increase in tobacco products can be considered as the 'population-wide' limit of the continuum and smoking cessation clinics the 'high-risk' limit. Then, mass media campaigns against smoking, work place smoking bans, and school based interventions can be ranked in-between the two extremes. The effectiveness and cost-effectiveness of population-wide prevention have been confirmed from multiple real world examples and modelling studies in multiple populations.[131–136]

Rose's approach to population-wide prevention was a paradigm shift among policy makers and researchers. However, some critics maintain that what he proposed more than three decades ago, may not apply today. They suggest that favourable downward trends in many risk factors (please refer to section 1.3 on page 17) have led the exposure of the population to the risk factors that Rose had considered to be much lower nowadays. Therefore, assuming an exposure limit below which no excess risk can be observed, the effectiveness of population-wide interventions is lower today than 30 years ago. Furthermore, high-risk prevention approaches have evolved over the years to use multivariate risk scores with lower threshold for treatments. Therefore currently, high-risk individuals are identified more accurately and the eligible for treatment population increases because of the lower risk threshold. Finally, the diffusion of the risk in the population is vital in Rose's approach. If the risk is densely clustered in specific population segments, high-risk strategies may be more effective.[137, 138]

Another area of critique for population-wide policies was their equity. Although Rose did not explicitly consider equity when proposed his taxonomy, Frohlich *et al.* suggested that population-wide prevention may increase inequalities. Their main argument is that some forms of population-wide interventions may not be equally effective across the population. For instance, health information campaigns against smoking have shown to be more effective among the more educated. Hence, they might increase smoking cessation rate among the more affluent and better educated and increase socioeconomic health inequalities. They called this the 'inequality paradox'.[139]

1.5.2 *The structural – agentic continuum*

In order to address the inequality paradox, McLaren *et al.* have further improved Rose's typology by making explicit some originally implicit assumptions about what constitutes population-wide prevention. Furthermore, they complemented Rose's typology by amal-

gamating advancements from the evolving field of health inequalities. The most important improvement was the introduction of another dimension to the classification of preventive interventions; the structural – agentic continuum.[140] The 'structure' and 'agent' notions originally stem from the field of social science.[141] In the public health policy context structural policies are those that "...promote health by altering the structural context within which health is produced and reproduced".[142] On the other hand, agentic policies are those that "...require mobilisation of an individual's resources, whether material or psychological".[143] Policies can fall anywhere in this continuum between these extremes. An increasing body of evidence suggests that structural preventive policies can potentially reduce socioeconomic health inequalities, while agentic policies are likely to increase them.[111, 112, 143–146]

For instance, let us consider three different policies to reduce excess salt consumption in the population. A population-wide structural policy could be mandatory reformulation of processed food to reduce their salt content. This would require no individual mobilisation to respond to the policy. On the other hand, a population-wide agentic policy could be a mass media campaign about the deleterious effect of excess salt consumption. This would require individuals to get exposed to the campaign, comprehend it, and act to change their behaviour/diet; socioeconomic gradient is likely in each of these stages. Finally, a high-risk agentic policy could use a nutritional questionnaire to identify those with excess salt consumption and then offer them dietary advice. This would require even higher individual mobilisation.

Apart from equity concerns, preventive policies with many agentic elements tend to be less effective due to attrition. Using the last example of the previous paragraph, some individuals may not respond to the questionnaire, may not attend the meeting to receive advice, may not comprehend the advice, and finally may not act upon the advice to modify their diet. The attrition is multiplicative and reduces dramatically the final effectiveness of the agentic policy.[146, 147]

1.5.3   *An alternative equity focused typology*

Benach *et al.* were inspired by Rose's and Graham's[148] typologies for prevention and they described a theoretical framework to classify public health preventive policies into four types, depending on their impact on different socioeconomic groups and their overall effectiveness on risk reduction. The first type includes policies that target only the most deprived. The second type includes universal policies across the whole population, with additional provisions for increased effectiveness among the most deprived. The third type includes policies that increase their effectiveness with the level of deprivation, but have no impact among the most affluent. Finally, the fourth type includes universal policies that increase their effectiveness with the level of deprivation, like the third type, but have an effect among the most affluent also.[149] Interestingly, Benach *et al.* used the term 'proportionate universalism' from Marmot to describe this fourth type. In the latest Marmot report about health inequalities in England, proportionate universalism policies were recommended to reduce the observed socioeconomic gradient in health.[98] In comparison,

policies targeting only the most deprived may be less effective on improving health overall, less effective on tackling inequalities, and less cost-effective.[98, 150]

1.5.4 *Intervention-generated inequalities*

From the previous descriptions of preventive typologies, it is evident that some preventive policies can in fact increase socioeconomic inequalities in health. This phenomenon is known as intervention-generated inequality and has been defined by White *et al.* as "unintended and unwanted variations in outcomes for individuals or population sub-groups that result from any element of any [health] intervention".[144] Unfortunately, many of the previous and current preventive efforts including health information campaigns and screening may be responsible for some of the currently observed health inequalities through intervention-generated inequality.[111, 112, 144, 151–154]

The Diderichsen model (section 1.4.2 on page 26) is a useful framework to conceptualise how some preventive interventions can generate inequalities. Some groups may have different exposure to the intervention. For example, a health advert in the Financial Times may be read mostly by managers and to a lesser extent by manual workers. Even if exposure is similar across the population, vulnerability to the intervention may be different. A television health advert may have different impact on viewers from different socioeconomic backgrounds. Finally, the differential consequences of the intervention arise from the differential resources an individual can mobilise to alter his/her behaviour.

The main advantage of any of these typology systems is that the effectiveness and equity of a preventive policy can be extrapolated from known aspects of the policy, even when direct empirical evidence is lacking. Nevertheless, the quantification of the effectiveness and equity of a policy requires knowledge of the distribution of risk in the population, the socioeconomic distribution of the population, the distribution of the policy impact, and their covariance. In addition, the element of time needs to be taken into account. Time trends in risk factor exposures and inequalities, lag times between exposure to risk factors and disease, and the diffusion time of a policy in the population are some examples highlighting the importance of time. In the next section I will describe how modelling can address some of these issues.

1.6 THE NEED FOR MODERN DECISION SUPPORT TOOLS

So far, I have briefly described how the interplay between patterns and trends in smoking, unhealthy diet, physical inactivity, and socioeconomic inequalities shape the patterns and trends in CVD, lung, and gastric cancer burdens and how preventive policies may influence this system. Public health policy makers need to make decisions in this complex dynamic system of exposures and outcomes based on often patchy evidence. In fact, public health policy makers have to lead system improvements in order to maximise health benefits for the population. They have to design, deliver, and evaluate evidence based public health

interventions that strike a balance between effectiveness, equity, and resource allocation, ideally while avoiding past mistakes.

Modelling can assist policy makers in the difficult task of making decisions about a volatile system. Weinstein *et al.* defined a model as "...a logical mathematical framework that permits the integration of facts and values and that links these data to outcomes that are of interest to health care decision makers".[155] In essence, models are abstractions of the perceived reality. The truth is that whenever a policy maker ventures a projection or simply thinks of a dynamic phenomenon (like NCDs trends or inequalities in health), he/she uses a model. An implicit, mental model that is based on the subjective perception that the policy maker has about the phenomenon. The assumptions of this implicit mental model are hidden; hence, they remain untested.

The huge advantage of modelling methodology is that it makes the implicit assumptions of mental models, explicit. Therefore, the explicit assumptions can now be critically appraised, calibrated to the data, and if possible validated. As knowledge progresses, some of the assumptions can be updated or replaced by newly acquired knowledge. In fact, explicit model assumptions can guide research efforts in order to fill knowledge gaps. The role of modelling in decision-making is to explain a phenomenon, illuminate its core dynamics, and highlight sources of uncertainty. As a result, more accurate predictions may arise. Yet, prediction is not the ultimate or most important goal of modelling. During the process of building and using a model, new questions worth asking may be discovered, new analogies with other phenomena may be revealed, and current knowledge may be challenged.[156]

Specifically in public health, since NCDs share common determinants (section 1.3 on page 17) preventive policies that target smoking, unhealthy diet, and physical inactivity will impact on more than one NCD. The joint prevention of NCDs had been observed in the past;[135, 157] however, it is only rarely explicitly considered in policy planning. Because of the different risk reversibility times, the policy impact time frame will be different for each disease. For example, an intervention that targets smoking will have an impact on CVD incidence within a couple of years, while the impact on lung cancer will be observed 10 to 20 years later.[135, 157] Additionally, because death is inevitable, decreasing the risk of a disease axiomatically increases the risk of any other disease over the life course. It is not feasible for traditional policy evaluation methods to capture the multiple, possibly competing, outcomes of a policy over long period of times.

Furthermore, traditional epidemiological research methods like cohort studies, randomised control trials, or systematic reviews, even when of outstanding quality, only rarely provide direct answers to public health policy makers. The 'evidence based medicine' doctrine and its 'evidence hierarchy' have served clinicians well for some well researched diseases. However, this doctrine cannot be directly applied in public health policy and practice, where questions are usually more complex and evidence more patchy. Petticrew *et al.* argued that in public health the full body of evidence may need to be used to inform different aspects of decision-making; different research designs answer different research questions. Therefore, the classic evidence based medicine evidence hierarchy may not apply in public health decision-making.[158] More recently, Mebius used a more philosophical approach to refute the notion of 'evidence hierarchy' and argued that "...when it

comes to evaluating the effectiveness of medical interventions, it is the evidence obtained from the methodology rather than the methodology that should establish the strength of the evidence".[159] Finally, Howick *et al.* maintained that mechanistic reasoning may be appropriate to guide decision-making when of high quality, and should not be vilified as happened in the past by evidence based medicine advocates.[160]

Modelling can address all the aforementioned concerns. Models synthesise information from multiple sources and use the full body of evidence. When good quality evidence is unavailable, explicit assumptions are based usually on mechanistic reasoning or experts' opinion. Then, model outputs are designed to give direct answers to policy makers, including an estimate of their uncertainty. The important role of NCD modelling in public health has been amalgamated by UK based NCD modellers in the 'Brighton declaration'.[161] Moreover, Smith *et al.* highlighted the importance of simulation modelling as a tool to design equitable policies and reduce health inequalities.[162] Yet, modelling approaches that attempt to dynamically simulate the complex system of how disease is generated in the system remain underutilised in epidemiology.[163, 164]

1.6.1 *Previously published NCD models*

Fone *et al.* conducted a systematic review on the use of modelling in population health and health care between 1980 and 1999. They identified 182 papers of which 120 had been published in the 1990s. The majority of models were about hospital organisation and scheduling, followed by models for cancer screening, and economic evaluation. Interestingly, they did not use a separate category for NCD models.[165] Three years later, Unal *et al.* focused their systematic review in policy models for CHD and identified 75 papers describing 42 models. Only six models had been published more than once.[166] More recently, Capewell *et al.* updated the systematic review by Unal *et al.* and expanded it to include the full CVD spectrum. They identified 70 CVD models and they concluded that transparency, comprehensiveness, and extended validation were lacking.[167] Finally, Speybroeck *et al.* included in their systematic review only models about socioeconomic inequalities in health. They identified 61 studies describing models using diverse designs and approaches.[168] All the reviews recognised the increasing use of modelling to explore a wide spectrum of research questions. However, they also identified the lack of objective methods to assess model quality and in many occasions the lack of transparency. Finally it was evident that the vast majority of models have only been published once and then they disappear from the literature.

During the first four months of this project, I endeavoured to a scoping review of existing public health policy models for NCDs. Due to lack of specific and widely used definitions of what consists 'public health policy' and what is a 'model' in this context, my search strategy was not sensitive and specific enough.[12] This issue had been identified and reported in the past by Fone *et al.*, who concluded that "...systematic reviews in the field of modelling should include reference list follow-up and contact with researchers in the

---

12 I tested the sensitivity and specificity of my search strategy by its ability to identify some prespecified 'key' modelling studies without producing an unmanageable number of hits, defined as more than 100 000.

field.".[165] Given that 1. recent, albeit less comprehensive, systematic reviews existed in the field and had been reported in the previous paragraph; and 2. a more comprehensive high quality systematic review would require substantial resources; jointly with my supervisors, we decided to focus my efforts on modelling. Therefore, I identified the models I describe below during my scoping review, after discussions with my supervisors and other modelling experts or by looking through the reference lists of other modelling papers. I only included models that I consider landmarks in NCD modelling for public health, and that were a personal inspiration to me not only because of their strengths but also because of their limitations. All the models (or modelling approaches in the case of comparative risk assessments) that I describe below, have been extensively used to inform policy in recent years, with the possible exception of Dynamo-HIA which is relatively new.

COMPARATIVE RISK ASSESSMENTS: One of the most widely used modelling methodologies in public health is comparative risk assessments. Comparative risk assessments use the notion of risk factor attributable disease burden, and they can estimate the hypothetical disease burden (usually assessed by mortality or incidence) under a counterfactual risk factor exposure distribution.

Many highly influential comparative risk assessments studies in the past, defined the counterfactual exposure distributions assuming a complete or near complete elimination of the respective risk factors[13].[20, 30, 32, 33, 170, 171] Although this approach is useful to quantify the overall disease burden attributable to the modelled risk factors and allow policy makers to prioritise their efforts targeting the risk factors with higher attribution to disease burden, the counterfactual exposure distributions are not directly linked to specific policy options. In other words, the impact of specific policies on the observed exposure distribution is not explicitly modelled. Therefore, this approach does not provide policy makers with all the necessary information about the effectiveness, equity, and efficiency of specific policy options.

Another limitation of comparative risk assessments is that they tend to be static by ignoring the time trends in exposures, disease burden, and demographics in the population. Finally, when mortality is used to assess disease burden, comparative risk assessments cannot account for competing mortality risks; the fact that since death is inevitable, decreasing mortality rate from a disease, increases the mortality rate from all other diseases, over the life course.

In the core of any comparative risk assessments there is an epidemiological formula, known as population attributable fraction (PAF). I will present and discuss the formula in the next chapter (section 2.4.1 on page 57 and section 2.9 on page 66). Many other epidemiological models use the same formula, some of which are described below, or in chapter 8 (section 8.2.1 on page 164). As I will describe in the next chapter, the model I propose also uses this formula. Although in a broad sense all these models can be considered as 'comparative risk assessment' models, they all offer ways to overcome some of the limit-

---

13 Murray and Lopez taxonomy of counterfactual exposure distributions includes four categories that correspond to the: theoretical minimum exposure; plausible minimum exposure, feasible minimum exposure; and cost-effective minimum exposure.[169]

ations of typical comparative risk assessments. For example, many models including the SimSmoke, DIETRON, and PRIME models, explicitly model the impact of specific policies to the counterfactual exposure distribution.[172–174]. Others, model two different time points and assume a linear time trend for exposures and disease burden.[175, 176]

SIMSMOKE: This is one of the most used models for public health policy. SimSmoke only includes smoking as a risk factor and models the mortality reductions from a set of tobacco control policies. As of 01/08/2016 Pubmed contained 36 papers that had used SimSmoke in multiple populations and have modelled several tobacco control policies, since 2000.[177] In terms of usefulness SimSmoke appears to be a successful model. However it is very specific in modelling tobacco control policies only. It does not model any socioeconomic health inequalities, and the output consists only of aggregated smoking related deaths (i. e. the policy effect cannot be stratified by specific smoking related disease mortality, nor morbidity).

CORONARY HEART DISEASE POLICY MODEL: This is probably the first model that was used to inform public health policy. Originally it was built specifically for the US setting but it has expanded to other countries recently. The Coronary Heart Disease Policy Model includes all major biological CHD risk factors and smoking and models CHD incidence, prevalence, and mortality. It models the impact of preventive policies on risk factors and through them on CHD burden. Additionally, to primary prevention the Coronary Heart Disease Policy Model also models CHD treatment options. As of 01/08/2016 Pubmed contained 27 published papers that were based on the model.[178] The main limitations of the model are that it is restricted to one disease only, it ignores important risk factors like unhealthy diet and physical inactivity, and it does not consider the wider determinants of health and socioeconomic health inequalities.

DYNAMO-HIA: This is a relatively new model that only appeared in the literature in 2012. However, since then only eight papers have been published based on it. DYNAMO-HIA has been primarily used to model the effect of primary prevention policies on health in a health impact assessment context for European populations. The authors offer the model as an application [14], as a generic modelling framework; however, the user interface leaves a lot to be desired. The expandable concept of DYNAMO-HIA allows the model to be useful for a wide range of preventive policies across Europe. Nevertheless, in the current implementation DYNAMO-HIA does not model the equity of these policies.[179]

IMPACT: The IMPACT family of models has been the most published and widely used for almost the last 15 years. More than 60 papers were based on IMPACT models, modelling populations from more than 20 different countries. The core ability of these models is that they attribute to known interventions, the observed difference in CHD mortality between two time points. The original incarnation was used to quantify the contribution of primary and secondary prevention to the observed CHD mortality decline, in a number of different

---

14 Available from http://www.dynamo-hia.eu/ free of charge.

countries and populations. Later incarnations were used to model the effectiveness of several primary prevention policies and more recently their equity.[175, 176, 180] The main limitations of the IMPACT models are that: 1. they are restricted to CHD only and do not include other diseases; 2. they only model CHD mortality (i.e. no incidence or prevalence); 3. they are static and ignore any changes between the two time points that are modelled; 4. they do not model the heterogeneity of the population and consequently, the heterogeneity is not propagated in the reported uncertainty intervals; and 5. they model population cohorts rather than individuals, which limits the type of policies that can be simulated (i.e. preventive policies targeting high-risk individuals (section 1.5.1 on page 31). I am describing later in the Methods chapter (chapter 2 on page 41) how my approach overcame all these limitations.

ARCHIMEDES: Unlike any other model that has been and will be described here, Archimedes comes from industry rather than academia and is a commercial product. About 15 papers have been published so far using the model, although given its commercial nature it is probably used by organisations for decision-making and planning without necessarily producing academic outputs. The impressive characteristic of the model is that it models human physiology and pathophysiology and how these evolve to manifest diabetes or CVD.[181] Its proprietary nature did not allow me to personally assess the model, and not many details are publicly available regarding the model internals. My impression is that Archimedes is mostly focused in pharmacological trials, and health care rather than public health settings. Therefore, modelling population-wide effectiveness or equity of primary prevention policies does not seem to be a priority.

POHEM: This is a very versatile model that was developed in 1994 and has been used since in at least 23 academic papers. It models a spectrum of NCDs including CVD and cancers. It can also model a wide spectrum of interventions both in public health and health care settings.[182, 183] Given that the model was originally developed in the 1990s it represents a huge achievement both in technical terms and in terms of usefulness and flexibility. Yet, the model can only be used for the Canadian population and the authors admitted that expanding POHEM to other populations would be almost impossible given the extensive data requirements. Moreover, although socioeconomic parameters like income, ethnicity, and education being included in the model, the equity of the modelled policies have not been explicitly estimated so far.

CISNET: Given the burden of cancer in the population, it is apparent that cancer modelling is underrepresented in the presented list of models. The National Cancer Institute in the US funds a consortium of researchers to develop cancer models, known as CISNET[15].[184] CISNET also maintains a very useful model registry that records available cancer models and their characteristics.[185] It is apparent from the registry that the majority of models focus on screening and treatment rather than primary prevention. It is

---

15 Stands for Cancer Intervention and Surveillance Modeling Network.

also evident that with the exception of SimSmoke each model models a specific cancer site, despite the common determinants of many cancers with other NCDs.

1.6.2 *Gaps in the modelling landscape*

LACK OF REUSABILITY: The models I presented in the previous paragraphs have been extensively used in the academic and policy debate. Unfortunately, they are the exceptions. In the current public health modelling landscape, most models are built to explore a very specific question in a very specific setting. For instance, hundreds of models appear in the literature only once and then they disappear completely. Therefore, the lack of reusability is evident and prohibits models to improve, evolve, and be assessed over time. In addition, the substantial resources that are needed to build a model are wasted and each new modelling attempt usually has to start from zero.

LACK OF TRANSPARENCY: The Brighton declaration identified that one of the challenges of NCD modelling is to successfully communicate the technical details and underlying assumptions of the model to technical and lay audiences.[161] I have to add that none of the models I described provide access to their source code. This means that the building elements of the model remain hidden from public scrutiny and any unintended errors in the code are less likely to be found. Proprietary code also forbids other researchers from improving and expanding an existing model forcing them to build a new one from scratch instead. Finally, the lack of transparency prohibits the dissemination of good modelling practices as they remain hidden and available only to authors.

In combination with the lack of reusability, lack of transparency increases dramatically the resources that are needed to apply modelling methodologies. This may discourage researchers and policy makers to opt for modelling support in the decision-making process. Even if a decision for modelling support is made, time and other resources are wasted to solve problems that have been already solved by other modellers. This may lead to lack of timeliness and relevancy of the model to the policy or academic debate.

LACK OF COMPREHENSIVENESS: While socioeconomic inequalities, risk factors, and NCD patterns and trends are interrelated it seems that almost none of the described models consider all of them simultaneously. Notable exceptions are the POHEM model which is limited to the Canadian population and some recent implementations of the IMPACT model which is limited to CHD mortality output only. Therefore, complex multicomponent policies are difficult to be modelled and fully assessed in the current modelling landscape.

All the issues above powerfully suggest the need for modern simulation based decision support tools in public health policy that are reusable, transparent, and take into account the complex dynamics of exposures and outcomes in the population.

1.7 AIMS AND OBJECTIVES

The primary aim of this project was to construct and validate a simulation model for public health policy that would be able to model any NCD (reusable), be open-source and avoid any proprietary code if possible (transparent), and consider the wider determinants of health and their interplay with risk factors and disease burden (comprehensive).

The equally important secondary aim was to use this model to integrate the available evidence and quantify the impact of existing and hypothetical counterfactual primary prevention policies on disease burden and health inequalities. Hence, to explore whether the estimated effectiveness and equity of these policies agree with the theoretical expectations according to their typology.

The objectives of this project were:

1. To build a model that has population dynamics, socioeconomic determinants, and risk factor trends as inputs; and CVD, lung cancer, and gastric cancer incidence, prevalence and mortality as outputs (chapter 2 on the facing page).

2. To extensively validate the model using external independent sources that ideally have not been used to inform model building (chapter 3 on page 77).

3. To use the model to quantify the contribution of statins in the observed decline of serum total cholesterol in England, between 1991 and 2011. (chapter 4 on page 103).

4. To use the model to quantify the effectiveness and equity of current primary prevention strategy for CVD in England (chapter 5 on page 121).

5. To use the model to quantify the effectiveness and equity of a spectrum of policies that have been applied in England since 2003 to reduce excess salt consumption in the population (chapter 6 on page 135)

6. To use the model to quantify the effectiveness and equity of hypothetical 'endgame' policies for smoking (chapter 7 on page 147)

METHODS

2.1 INTRODUCTION

In the previous chapter I argued the necessity of reusable, transparent, and comprehensive modelling approaches to inform public health policy. In this chapter I will describe my approach to build a simulation model that fulfils those criteria. The modelling methodologies spectrum is complex, evolving, and diverse; thus, hard to be classified into specific subgroups. Several typologies have been proposed so far, but none is without limitations nor widely accepted outside the scope it was proposed for. One of the most useful modelling taxonomies in public health decision-making appears to be the one proposed by the International Society for Pharmacoeconomics and Outcomes Research jointly convened with the Society for Medical Decision Making.[186] They have identified three large families of model structures: state transition models, discrete event simulation, and dynamic transition models. In the following paragraphs, I will briefly present the three families and I will justify my decision to build a state transition model.

STATE TRANSITION MODELS: State transition models are preferred when the decision problem can be described by well defined health 'states'. Population cohorts or individuals are then modelled to populate these states and are allowed to move in different states based on pre-specified 'transition probabilities'. The modelled health states need to be defined in a way that are mutually exclusive and collectively exhaustive and cohorts or individuals are allowed to be in only one state at a given time. Depending on whether the model simulates cohorts or individuals, it is called macro- or microsimulation respectively[16].[187]

DISCRETE EVENT SIMULATION: Discrete event simulation is preferred when modelling of a complex system is necessary or when time management[17] is important. The core structural units in discrete event simulation are the 'entities'. Entities are objects that usually represent individuals and have a set of 'attributes'; for instance, age, sex, medical history etc. Attributes can be manipulated dynamically during the simulation and determine how an entity reacts to an 'event'. An event is simply anything can happen to the entity or the environment during the simulation. Depending on the nature of the research question, the simulation can be constrained by the introduction of scarce 'resources' to the simulated system. The entities then compete to utilise the resources and form 'queues' waiting for the resource to be available.[188]

---

16 Macrosimulations are also known as 'Markov models'.
17 For example, to model waiting times of patients awaiting a planned medical procedure.

DYNAMIC TRANSMISSION MODELLING: Dynamic transmission modelling is mostly relevant in the field of communicable diseases. Its core idea is that the probability of new infection is directly related to the proportion of those already infected in the population.[189]

2.1.1   *Conceptualising the problem*

The conceptualisation of the problem at hand is one of the first and most important steps in the modelling process.[190] My aim was to model the equity and effectiveness of primary prevention policies for multiple NCDs (section 1.7 on page 40). Decisions about time management or resources allocation were not relevant in this case. Therefore, I chose a state transition approach as the most appropriate for my aims. Modelling equity, multiple risk factors, and multiple diseases in a competing risk framework requires a level of complexity not easily manageable in a macrosimulation framework because of the large number of states. Furthermore, modelling policies that target high-risk individuals (section 1.5.1 on page 31) in the population require these individuals to be modelled in the simulation. In fact, from the model I described in the previous chapter (section 1.6.1 on page 35), the only model that appears to have similar capabilities is POHEM (section 1.6.1 on page 38), which is a microsimulation. Hence, I opted for a microsimulation approach for my model.

2.1.2   *Definition, history, and typology of microsimulations*

The International Microsimulation Association defines microsimulation as "...a modelling technique that operates at the level of individual units such as persons, households, vehicles or firms. Within the model each unit is represented by a record containing a unique identifier and a set of associated attributes – e.g. a list of persons with known age, sex, marital and employment status. [...] A set of rules (transition probabilities) are then applied to these units leading to simulated changes in state and behaviour. These rules may be deterministic (probability = 1), such as changes in tax liability resulting from changes in tax regulations, or stochastic (probability ≤ 1), such as chance of dying, marrying, giving birth or moving within a given time period. In either case the result is an estimate of the outcomes of applying these rules, possibly over many time steps, including both total overall aggregate change and, crucially, the distributional nature of any change. Given the emphasis on changes in distribution, microsimulation models are often used to investigate the impacts on social equity of fiscal and demographic changes (and their interactions)".[191]

Microsimulation was first proposed by Guy Orcutt in a 1957 seminal paper, as a new modelling methodology for econometrics to analyse and forecast the individual impacts of economic and social policies.[192] In practice, microsimulations use real world information to create a synthetic world, where virtual experiments can be performed.[193] The method only gained popularity four decades after its original inception due to its large data and computational requirements. Nowadays, microsimulation is a well accepted method to support decision-making in health care policy, although its use in public health policy appears limited.[194–196]

Zucchelli *et al.* summarised a typology for microsimulations based on previous proposals. Its advantage is that it is practical and applicable in the context of health policies evaluation. The typology is based on two dichotomies: arithmetical versus behavioural models, and static versus dynamic models.[196]

ARITHMETICAL VERSUS BEHAVIOURAL MODELS: Arithmetical microsimulations ignore individual behavioural response to the modelled policy change. For example, a mandatory reformulation of processed foods that would lead to a fast decline in their salt content, may drive consumers to change their behaviour as a response, and add discretionary salt before consumption. An arithmetical microsimulation would not simulate the behavioural response to policy change; on the contrary, a behavioural microsimulation would do that. From a practical perspective, the choice between the two is more on the user side and what assumptions the user feels necessary to include in the simulated scenario, rather than a fundamental model design choice.

STATIC VERSUS DYNAMIC MODELS: Unlike the previous categorisation, the differentiation between static and dynamic microsimulations is a fundamental one. Static microsimulations do not include the element of time. Their analysis is restricted to a single point in time or a set of points in time, ignoring all the intermediate time points. For example, consider a microsimulation that explores the effect of a newly introduced lipid lowering medication to the population. A static microsimulation would only examine the two putative states for each individual; before and after the introduction of the new medication.

By contrast, dynamic microsimulations simulate individual life courses over time. The attributes of the synthetic individuals are updated in each time interval. Therefore, ageing, population trends, and policy effects can be simulated dynamically. Dynamic microsimulations may consider time as a discrete or continuous variable. Discrete and continuous time microsimulations have vital differences in their conception, implementation, and core assumptions that are beyond the scope of this thesis.

2.2 HIGH LEVEL DESCRIPTION OF IMPACT_{NCD}

IMPACT_{NCD} is a discrete time, dynamic, stochastic microsimulation. Within IMPACT_{NCD} each unit is a synthetic individual and is represented by a record containing a unique identifier and a set of associated attributes. Age, sex, quintile groups of Index of Multiple Deprivation (QIMD)[18], salt consumption, BMI, SBP, total serum cholesterol, diabetes mel-

---

18 QIMD is a measure of relative area deprivation based on the Index of Multiple Deprivation.[197] According to this system, all Lower Super Output Areas in England (average population of 1500) are ranked in order of increasing deprivation, based on seven domains of deprivation: income; employment; health deprivation and disability; education, skills and training; barriers to housing and services; crime and disorder, and living environment. For the ranking, individual level information about the habitats of these areas is used from multiple sources. Then, the QIMD is formed from the quintiles of the above index, one through five, where quintile one is considered the 'least deprived' and quintile five the 'most deprived'.

litus[19](binary variable), smoking status (current / ex- / never-smoker), smoking history (duration measured in years, and intensity measured in cigarettes per day), environmental tobacco exposure (binary variable), fruit and vegetable consumption and physical activity as the set of associated attributes. A set of stochastic rules are then applied to these individuals, such as the probability of developing CHD or dying, as the simulation advances in discrete annual steps. The output is an estimate of the burden of CHD, stroke, lung cancer, and gastric cancer in the synthetic population including both total aggregate change and, more importantly, the distributional nature of the change. This allows for an investigation of the impact of different modelled policies on social equity.

IMPACT$_{NCD}$ is a complex model that simulates the life course of synthetic individuals under counterfactual scenarios and currently consists of two modules: the 'population' module and the 'disease' module. Figure 2.1 on the facing page highlights the steps of the algorithm that generate the life course of each synthetic individual. Step 1 only runs at the beginning of each simulation. Following steps 2-7 are calculated annually (in simulation time) for each synthetic individual until the simulation horizon is reached, or death occurs. I will fully describe IMPACT$_{NCD}$ by describing the processes in each of these steps. The description is from an epidemiological perspective and focuses on the processes rather than the technical details. Table 2.2 on page 70 summarises the data sources that have been used to inform the parameters of IMPACT$_{NCD}$, and table 2.1 on page 69 presents its main assumptions and limitations.

To ensure transparency, the source code and all parameter input files are available in https://github.com/ChristK/IMPACTncd/tree/Thesis_model_version under the GNU GPLv3 licence. The GNU GPLv3 licence is a free, copyleft license for software and other kinds of works (https://www.gnu.org/licenses/gpl-3.0.en.html). The licence grants the right for everyone to use the model as it is or modify it, although if someone releases a modified version of the model to the public, the licence requires the modified source code to be publicly released under the GNU GPLv3 licence as well; therefore, to remain open-source.

---

19 I defined as diabetics those with self-reported medically diagnosed diabetes or glycated haemoglobin ≥ 6.5 %, excluding pregnancy only diabetes.
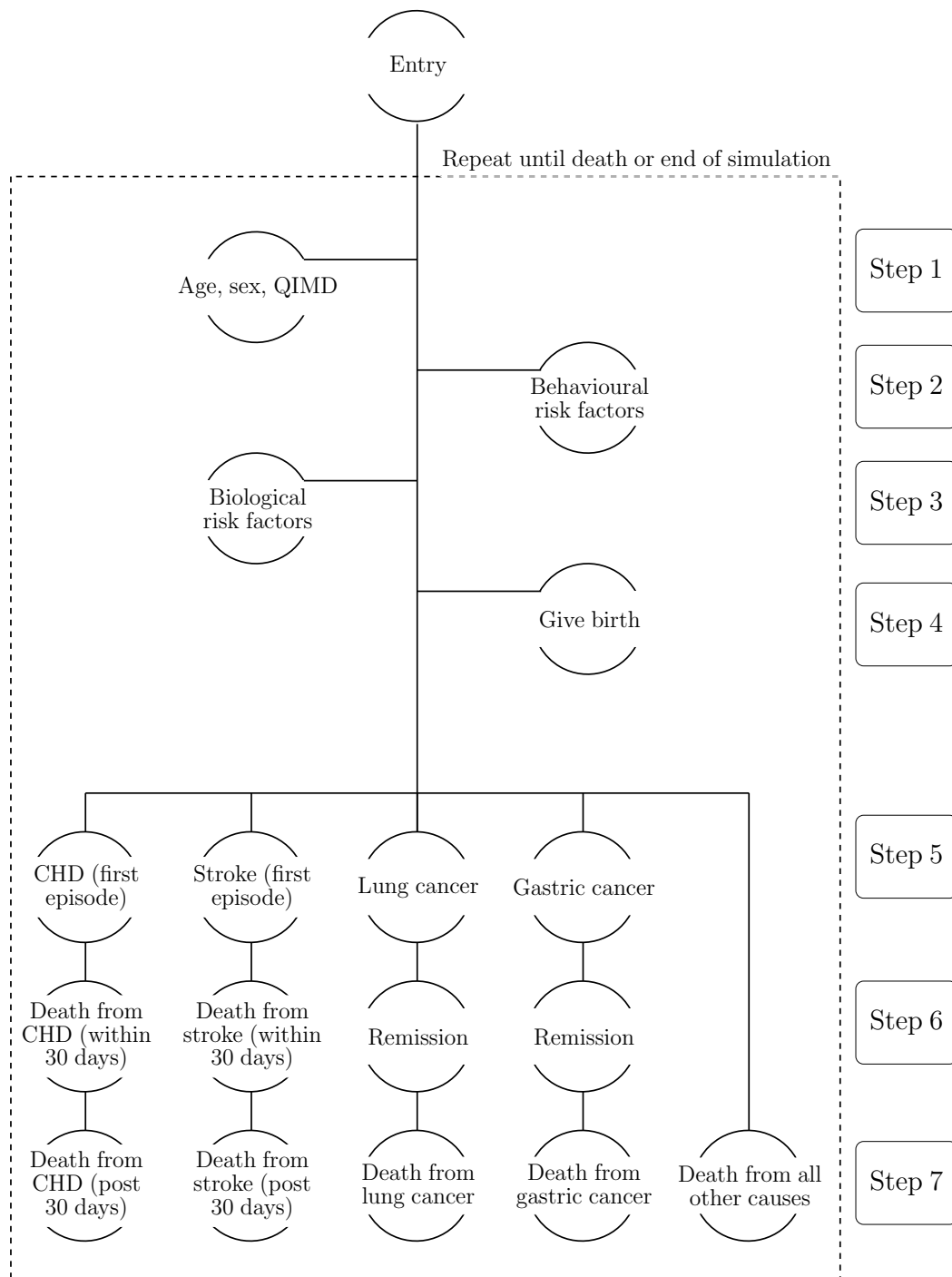
Figure 2.1: Simplified IMPACT$_{NCD}$ algorithm for individuals. For each step, the algorithm uses information from all appropriate previous steps.

The 'population' module consists of steps 1 to 4 in figure 2.1 on page 45. Synthetic individuals enter into the simulation in the initial year. The number of synthetic individuals that enter into the simulation is user-defined and depends on the rarity of the simulated diseases in the population. The characteristics of the synthetic individuals were informed from the HSE, and the algorithm ensures that the age, sex, and QIMD distribution of the sample is similar to the age, sex, and QIMD distribution of the English population in the initial year.

2.3.1   *Health Survey for England profile*

HSE consists of a series of cross-sectional health surveys that have been conducted annually since 1991 and are representative of the community dwelling population in England. While the focus each year was in a different aspect of population health, the core questions have remained relatively stable over the years. Hence, secular trends of risk factor exposures can be extracted from the survey. HSE has a complex sample design and selection bias weighting has been applied since 2003 to adjust for non-responders. The response rate has dropped from around 70 % in the 1990s to around 60 % in more recent years. The participants are first interviewed and then if they consent, a nurse visits them to conduct further measurements and collect biological samples. More information about the series can be found elsewhere.[198]

2.3.2   *Estimating exposure to risk factors*

In steps 2 and 3 (figure 2.1 on page 45), IMPACT$_{NCD}$ estimates the exposure of each synthetic individual to the modelled risk factors. It is essential that the risk profile of the synthetic individuals should be similar to the risk profiles that have been observed in the actual English population. For this, I first built a static 'close to reality' synthetic population of England, from which I sample the synthetic individuals for every new run of the simulation. Then, I simulate individual exposure trajectories for all synthetic individuals to generate individual life courses. In the following paragraphs I will describe the processes to achieve this.

2.3.2.1   *Generating the 'close to reality' synthetic population for IMPACT$_{NCD}$*

Synthetic populations are a vital component of microsimulation models. Research and innovation in synthetic populations methodology appears to be more active in the field of transportation science.[199] The use of synthetic populations in epidemiology is limited and mostly in the field of infectious diseases epidemiology.[200] In my implementation, the 'close to reality' synthetic population ensures that the sample of synthetic individuals for the simulation is drawn from a synthetic population similar to the real one in terms of the correlation structure for age, sex, socioeconomic circumstance, and risk factor exposures.

For this, I have adapted the statistical framework originally developed by Alfons *et al.* in order to better reflect epidemiological principles.[201]

Briefly, Alfons *et al.* method uses a nationally representative survey of the population to generate a 'close to reality' synthetic population. Therefore, the method expands the often small sample of the survey into a significantly larger synthetic population, while it preserves the statistical properties and respects the correlation structure of the original survey. This is possible by fitting multinomial regression models to the survey data and then sample the synthetic individuals from the respective conditional distributions. Practical applications and examples of the original method can be found elsewhere.[202, 203]

Traditional methods for population synthesis use iterative proportional fitting or combinatorial optimisation to derive the synthetic population from a given survey.[199] The main advantages of the Alfons *et al.* approach over these alternatives are that: 1. it takes into account the hierarchical structure of the sample design of the original survey (i. e. individuals within households, within larger geographical areas); and 2. it can generate trait combinations which were not present in the original survey but are likely to occur in the real population. The latter is particularly important, because it avoids bias from excessive repetition of a limited set of trait combinations present in the original survey sample. For example, the original survey may only have two 80 year old male participants, both ex-smokers. Unlike other methodologies, the approach proposed by Alfons *et al.* can generate 80 year old male synthetic individuals, who are never or current smokers despite the fact that these combinations of traits were not present in the survey. It also prohibits extreme outliers present in the original survey to be overrepresented in the synthetic population.

My approach in synthetic population generation consists of four stages from which the first is common with the original method proposed by Alfons *et al.* The following stages have been adapted in order to reflect ideas and principles from the 'wider determinants of health' framework (section 1.4.1 on page 24) and the 'social production of disease' model (section 1.4.2 on page 26). The main notion is that upstream factors such as the socioeconomic conditions, influence individual behavioural risk factors (e. g. diet, smoking, and physical activity), which in turn influence individual downstream risk factors such as BMI, SBP, and total cholesterol. The four stages are:

1. Set up of the household structure.

2. Generate the socioeconomic variables of the synthetic individuals.

3. Generate the behavioural variables of the synthetic individuals.

4. Generate the biological variables of the synthetic individuals.

Each stage is informed by relevant information from all previous stages. This is performed in the spirit of the original method by fitting a series of multinomial regression models to the survey data that use predictors from previous stages, and then use the models to predict the traits of the synthetic individuals. The output is a static synthetic population that is a 'snapshot' of the real population at a specific moment in time.

For the purpose of this thesis I produced two synthetic populations representing two snapshots of the English population, in years 2006 and 2011. All the variables of the 2011

synthetic population were informed by the HSE2011 except physical activity which was informed by HSE2012.[204, 205] For the snapshot of the population in 2006, I used three consecutive surveys from years 2005, 2006, and 2007;[206–208] a decision that balances the benefits of the decreased sampling error against the introduced bias from the secular trends of risk factors. The R language for statistical computing and the R packages 'simPopulation' and 'simPop' were used to implement the method[20].[209–211] A more detailed description of the four stages follows, based on the 2006 synthetic population.[21]

STAGE 1: HOUSEHOLD STRUCTURE   The household size, the age, and sex of the individuals in each household that have been recorded in HSE were used to inform the synthetic population, stratified by Strategic Health Authority[22]. Strategic Health Authority was the only variable with spatial information in HSE and it was used as a proxy, to include some spatial information to the synthetic population.

STAGE 2: SOCIOECONOMIC VARIABLES   Once the basic age, sex, household, and spatial information of the synthetic population was generated, other socioeconomic information was built up. The area deprivation of each household was generated dependent on the age and sex of the head of the household, stratified by Strategic Health Authority. QIMD was used as a measure of relative area deprivation. Then, the equivalised income quintile groups[212] of each household was generated, dependent on the age and sex, stratified by QIMD. Finally, the employment status of the head of the household was generated using the National Statistics Socio-Economic Classification[213], dependent on age and sex, stratified by QIMD.

STAGE 3: BEHAVIOURAL VARIABLES   In this stage, the behavioural variables of each synthetic individual were predicted using information that was generated in the previous stages. Portions of fruit and vegetable consumed per day, days achieving more than 30 min of moderate or vigorous physical activity per week, smoking status and smoking histories were generated, dependent on age, sex and stratified by QIMD. For smoking histories, smoking duration and intensity for active smokers, and smoking duration, intensity, and years since cessation for ex-smokers were also generated in this step. To propagate the correlation structure between smoking duration and intensity to the synthetic population, I first estimated smoking intensity dependent on age, sex and stratified by QIMD. Then I used smoking duration as a dependent variable together with age and sex and stratified by QIMD to predict intensity and years since cessation, where applicable. This correlation is important when both duration and intensity are used to calculate the cumulative risk of smoking for lung cancer. The exposure to environmental tobacco smoking was also predicted in this stage dependent on age, sex, smoking status and stratified by QIMD.

---

20 The R scripts are available at https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/SynthPop/Synthetic%20Population%20UOM%20050607.R and https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/SynthPop/Synthetic%20Population%20UOM%202011.R.

21 A similar process was followed for the 2011 synthetic population

22 Strategic Health Authorities were ten large geographic areas, part of the structure of the NHS in England before 2013.

Once again the correlation between smoking status and exposure to environmental tobacco smoking is important because the risk of environmental tobacco smoking for active smokers is considered negligible. Moreover, the use of statins and the use of antihypertensive medication (two separate binary variables) were also predicted dependent on age, sex and stratified by QIMD. Finally, a preliminary salt consumption estimation was performed during this stage. HSE collected spot urine sodium measurements, which are less reliable to 24 h urine sodium measurements.[214, 215] To overcome this limitation IMPACT$_{NCD}$ adds another processing layer for the final estimation of salt consumption to integrate information from other data sources that used 24 h urine collection. This is described separately in section 2.3.2.4 on page 51.

STAGE 4: BIOLOGICAL VARIABLES   The final stage is the prediction of the biological variables of the synthetic individual. Widely accepted causal pathways that have been observed in cohort studies have been used to identify associations between biological and behavioural variables and how the risk of some risk factors is partly mediated through other risk factors. The WHO 'Comparative quantification of health risks' report was primarily used as a guide to identify causal pathways and has been described in section 1.3 on page 17.[30] The aim was to capture the correlation structure of these related risk factors in the HSE and propagate it to the synthetic population.

BMI mediates part of its risk for CVD through SBP, total serum cholesterol, and diabetes mellitus.[216–221] Thus, BMI was the first to be predicted in the synthetic population dependent on age, sex, and fruit and vegetable consumption, stratified by QIMD. Fruit and vegetable consumption was used as a proxy to healthy diet and this improved the predictive power of the model.[222] On the contrary, while physical activity is related to BMI in my case the inclusion of physical activity as a predictor weakened the predictive power of the model[222–224]; hence, I excluded physical activity from the predictors. The cross-sectional nature of the data and the associated reverse causality bias may have played a role in this. The numerical instability of the model from the inclusion of another categorical variable that led to further segmentation of the relatively small sample could be another plausible explanation.

After BMI, diagnosed diabetes mellitus status was predicted for the synthetic individuals dependent on age, sex, QIMD, and stratified by BMI deciles. Then undiagnosed diabetes mellitus was predicted using the same predictors. I defined undiagnosed diabetes similarly to the HSE definition, as having glycated haemoglobin greater than 6.5 % and not reporting diagnosed diabetes mellitus in the survey questionnaire. Pregnancy related diabetes mellitus was excluded from both models. Afterwards, total serum cholesterol was predicted dependent on age, sex, BMI, statin use, and fruit and vegetable consumption, stratified by QIMD. Finally, SBP was predicted dependent on age, sex, BMI, smoking status, and salt consumption, stratified by QIMD. The association of smoking with hypertension has been observed in longitudinal studies.[37, 225, 226] It is worth noting here that the use of QIMD for stratification in behavioural and biological risk factors prediction, allows for possible interaction between socioeconomic and behavioural variables in the prediction of biological risk factors.

SMOOTHING FOR CATEGORICAL VARIABLES   Despite the use of three consecutive years of HSE to prime the synthetic population, there were still a small number of elderly participants, when stratified by sex and QIMD. This created large troughs and peaks in the prevalence of smoking and diabetes mellitus for certain age groups that could have been propagated to and biased the synthetic population. To avoid the issue I applied smoothing to HSE smoking and diabetes data. This was performed first by constructing the contingency table of smoking status by age and sex. Then I used local regression to smooth the numbers of participants by age, stratified by sex and smoking status.[227] Finally, I used the smoothed contingency table to recalibrate the survey weights using the raking method.[228, 229] I followed the same approach for diabetes mellitus.

The final output of the process overall was to create a snapshot of the 2006 English population as it was captured by the HSE. The synthetic population has similar statistical properties to the HSE sample that has been used to prime it. More importantly, the synthetic population models both the 'differential exposure' and the 'differential vulnerability' of the 'social production of disease' theoretical model (section 1.4.2 on page 26). The approximately 55 million synthetic individuals with a combination of traits similar to the community dwelling population that have been generated, define the initial state of the population during the microsimulation.

Next, I will describe how the exposures of the synthetic individuals can be updated dynamically as the microsimulation advances through time. The synthetic population has been validated against the original HSE sample in the validation chapter (chapter 3 on page 77).

2.3.2.2   *IMPACT$_{NCD}$ implementation for life histories simulation*

IMPACT$_{NCD}$ only applies the previous process for the initial year of the simulation. As the simulation evolves, all variables are recalculated to take into account age and period effects. This feature justifies the classification of IMPACT$_{NCD}$ as a dynamic microsimulation. The exact method depends on the nature of each variable and the available information. Generally, I fitted appropriate statistical models to the HSE data for years 2001 to 2012 to capture the secular trends of risk factors by age, sex, and QIMD. I used the R package 'survey' to handle all HSE data and fit the models to account for their complex sampling design.[230, 231] During the simulation IMPACT$_{NCD}$ projects the models into the future and use them to make predictions about the synthetic individuals.[204–208, 232–238]

2.3.2.3   *Age, sex and socioeconomic variables*

As the simulation advances in annual circles the age of the synthetic individuals in the model increases by one year. The sex and socioeconomic variables remain unaltered throughout the simulation. Therefore, social mobility is not simulated in the current version of IMPACT$_{NCD}$.

2.3.2.4   *Salt*

As I have already mentioned in section 2.3.2.1 on page 48, HSE measured sodium excretion from spot urine. The obvious next step would be to use the INTERSALT equation to approximate daily sodium consumption from spot urine.[219, 239] However, while the INTERSALT equation is an acceptable method to estimate the mean sodium consumption of the population, it tends to overestimate low measurements and underestimate high measurements, compared to the gold standard of sodium estimation from 24 h urine collection. Therefore, it distorts the distribution of salt consumption in the population and would introduce unnecessary bias.[214, 215]

HSE was not the only source of information regarding the exposure of population to salt. Between 2001 and 2011 four nationally representative surveys of the population were contacted specifically to investigate population exposure to salt.[68, 240–242] These surveys used 24 h urine collections to estimate sodium excretion; hence, salt consumption.[23] Unfortunately, the reported results from the aforementioned sodium surveys were aggregated, stratified by age group and sex. Individual level primary data were not available to researchers. Hence, the sodium surveys results could not directly inform the synthetic population using the method described in section 2.3.2.1 on page 46. Hence, I developed a stochastic process that integrates the individual level information from the HSE with the less flexible but more accurate information from the sodium surveys (please refer to appendix A.1 on page 183 for a detailed description). The main advantage of this approach is that it uses all the available information from the 24 h urine sodium surveys, while augmenting it with information regarding socioeconomic gradients and correlation with other risk factors and especially SBP, from spot urine measurements. The stochastic nature of the process allows its uncertainty to be estimated with Monte Carlo methods and is included in the reported uncertainty interval (UI).

2.3.2.5   *Fruit and vegetable consumption and physical activity*

Both fruit and vegetable consumption (portions per day) and physical activity (days with more than 30 min of moderate or vigorous activity per week) were modelled as ordinal variables. A proportional odds logistic regression model was fitted in HSE individual level data with fruit and vegetable portions as the dependent variable and year, second-degree polynomial of age, sex, QIMD and their interactions as the independent variables. Similarly, for physical activity a similar model was fitted in the HSE2006, HSE2008 and HSE2012 data. These models are used for individual level predictions about the synthetic individuals as the simulation evolves.[24]

---

23 I assume that all consumed salt is excreted through urine and all the sodium that is excreted in urine comes from the consumed salt. This is a common assumption in epidemiological literature

24 The R objects for the models are available at https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/Lagtimes/fv.svylr.rda and https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/Lagtimes/pa.svylr.rda.

2.3.2.6 *Smoking*

The 'close to reality' synthetic population is an accurate snapshot of active, ex-, and never-smokers in 2006, as it was captured in HSE. IMPACT$_{NCD}$ can use one of two different methods to simulate smoking histories, depending on scenario specification. The first method uses transitional probabilities for smoking initiation, cessation, and relapse to generate and record smoking histories for the synthetic individuals. For smoking initiation and cessation probabilities, logistic regression models were fitted in HSE data with age, sex, and QIMD as the independent variables. A similar approach was followed for relapse probabilities with years since cessation, sex, and QIMD as the independent variables. This method allows policies to be simulated using their impact on smoking initiation, cessation, and relapse probabilities.[25]

The second method uses two binomial regression models, one to estimate the prevalence of ever-smokers in the population, and one to estimate the prevalence of active smokers among ever-smokers; both by age, sex, and QIMD. As before, the binomial models were fitted in the HSE data and then are used for projections. The advantage of this method is that simulation scenarios can be specified using smoking prevalence instead of transitional probabilities for smoking initiation, cessation, and relapse. The disadvantage is that some ever-smokers may need to become never-smokers during the simulation to comply with the predicted prevalence. To minimise the bias, IMPACT$_{NCD}$ selects the ex-smokers with the longest abstinence periods to be reclassified as never-smokers. This method allows policies to be simulated using their impact on smoking prevalence.[26]

Independent of the method used to estimate the smoking status of each synthetic individual, IMPACT$_{NCD}$ tracks smoking duration for ever-smokers, and years since smoking cessation for ex-smokers. It also estimates smoking intensity (in cigarettes per day) for every active smoker in each simulated year. For this, I fitted a quasi Poisson regression model in HSE data with year, age, sex, and QIMD as the independent variables.[27] IMPACT$_{NCD}$ tracks mean smoking intensity over the last 10 years for each active smoker and uses this and smoking duration for further estimation of the cumulative risk of smoking for cancers.

2.3.2.7 *Environmental tobacco smoking*

For environmental tobacco smoking exposure I assumed a linear relation with smoking prevalence, stratified by QIMD. I assumed no intercept; therefore, when smoking prevalence reaches zero, environmental tobacco smoking prevalence will be zero too. With this

---

25 The R objects for the models are available at https://github.com/ChristK/IMPACTncd/blob/Evaluation_of_ UK_salt_strategy/Lagtimes/smok.start.svylr.rda, https://github.com/ChristK/IMPACTncd/blob/Evaluation_o f_UK_salt_strategy/Lagtimes/smok.cess.svylr.rda, and https://github.com/ChristK/IMPACTncd/blob/Thesis _model_version/Lagtimes/smok.cess.success.parabola.rda.

26 The R objects for the models are available at https://github.com/ChristK/IMPACTncd/blob/Thesis_model_ver sion/Lagtimes/smok.start.svylr.rda and https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version /Lagtimes/smok.cess.svylr.rda.

27 The R objects for the model is available at https://github.com/ChristK/IMPACTncd/blob/Thesis_model_versi on/Lagtimes/cigdyal.svylr.rda.

approach, IMPACT$_{\text{NCD}}$ draws a number of synthetic individuals and assigns them as exposed to environmental tobacco smoking in order to comply with the predicted prevalence of environmental tobacco smoking for each QIMD and simulated year.

2.3.2.8 *Continuous biological risk factors*

In IMPACT$_{\text{NCD}}$, the value of each continuous biological risk factor (BMI, SBP, total cholesterol) is calculated in a two stage process for each synthetic individual and simulated year. The first stage simulates ageing effects, while the second stage simulates period effects. I followed this approach mainly for two reasons. First, to simulate physiological mechanisms of ageing; for example the change of lipid profile in post-menopausal women, or the increase of SBP due to age related stiffening of the arteries. Second, because the variance of continuous biological risk factor distributions increases with age, I had to model this increase of variance with age to minimise bias. In the following paragraphs I describe the two stages:

STAGE 1:   IMPACT$_{\text{NCD}}$ tracks the percentile ranks of the biological risk factors of the synthetic individuals by 5-year age group, sex, and QIMD. These percentile ranks remain fixed for each synthetic individual throughout the simulation. The percentile ranks are translated back to risk factor values by matching them with the percentile ranks from a sample of the initial synthetic population of same age group, sex, and QIMD. For example, consider a 20-year old male synthetic individual living in a QIMD 3 area, and having SBP of 130 mmHg in 2006. Let us assume that his SBP corresponds to a percentile rank of 0.70 for his age group, sex, and QIMD. To estimate his SBP fifty years later, in 2056, when the same synthetic individual will be 70-year old, IMPACT$_{\text{NCD}}$ assumes that his percentile rank remained stable. Then, IMPACT$_{\text{NCD}}$ finds a 70-year old synthetic individual living in a QIMD 3 area in 2006 having the same percentile rank for SBP (0.70), and assigns his SBP (let us assume 146 mmHg) to the first synthetic individual. Figure 2.2 on the following page illustrates the previous example. Despite individuals retaining their percentile ranks for the respective risk factors throughout the simulation this stage remains stochastic. Every time a percentile rank is translated to a risk value a different sample from the initial synthetic population is drawn. Therefore, the same percentile rank is translated to a slightly different value of risk factor.

STAGE 2:   Risk factor values estimated from the previous stage ignore any period effects of risk factors. Period effects are implemented in this stage. Similarly to the approach followed for other variables, I fitted regression models to the HSE data. For the BMI model, year, age, sex, QIMD, and physical activity were the independent variables. For the SBP model, year, age, sex, QIMD, smoking status, BMI, and physical activity were the independent variables. Finally, for the total cholesterol model, year, age, sex, QIMD, BMI, and fruit and vegetable consumption were the independent variables. The independent variables
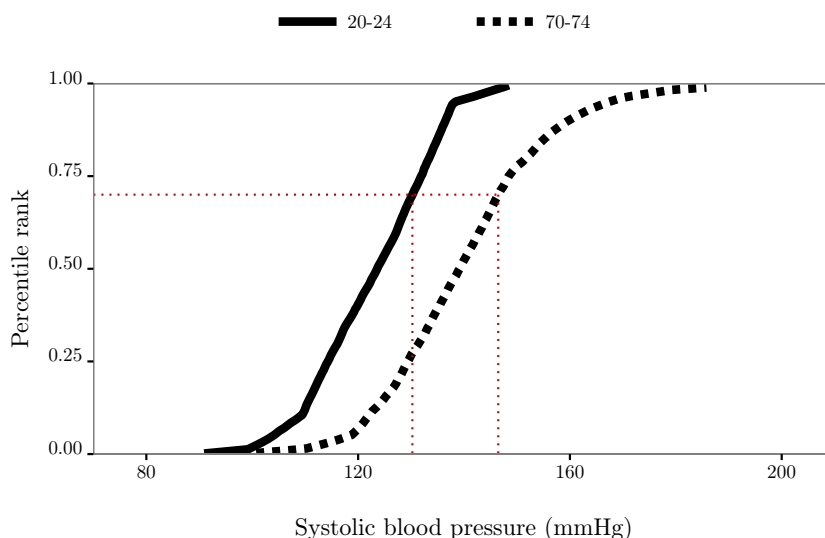
Figure 2.2: Plot of the systolic blood pressure against its percentile rank for male synthetic individuals in age groups 20 to 24 and 70 to 74, living in quintile groups of Index of Multiple Deprivation 3 area. The red dotted lines highlight how the same percentile rank corresponds to different blood pressure value for different ages.

for each risk factor were selected based on known associations from longitudinal studies as in section 2.3.2.1 on page 49.[28]

IMPACT$_{NCD}$ estimates the final values of the respective risk factor that will be used for risk estimation by, first, calculating the distance from the mean for each risk factor value stratified by 5-year age group, sex, and QIMD. For instance, if a synthetic individual has SBP of 140 mmHg and the mean SBP in the respective group of same age group, sex, and QIMD is 130 mmHg, the distance from the mean is 140 – 130 = 10 mmHg. Then, IMPACT$_{NCD}$ uses the models to predict the new mean of the respective risk factor by year, age group, sex, and QIMD and add it to the calculated distances. Hence, the estimated values for the continuous biological risk factors take into account both ageing and period effects.

2.3.2.9 Diabetes mellitus

The last risk factor to be simulated for life histories is diabetes mellitus. As with smoking, the 'close to reality' synthetic population is an accurate snapshot of diagnosed and non-diagnosed diabetics in 2006. IMPACT$_{NCD}$ uses the validated for the English population Qdiabetes[29] algorithm to calculate annual transitional probabilities of non-diabetic synthetic individuals of developing diabetes mellitus.[243, 244] I assumed diabetes mellitus

---

28 The R objects for the models are available from https://github.com/ChristK/IMPACTncd/blob/Thesis_model _version/Lagtimes/bmi.svylm.rda, https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/Lag times/sbp.svylm.rda, and https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/Lagtimes/chol .svylm.rda.

29 Previously known as QDscore.

is an incurable chronic condition; therefore, synthetic individuals who develop diabetes mellitus remain diabetics through their life course. In reality, the number of patients with diabetes who truly become and remain normoglycaemic is very small.

2.3.3 *Exposure and disease lag times*

In section 1.3 on page 17 I described that risk factors have different reversibility lag times for different diseases. This means that past level of exposures defines current disease risk. $\text{IMPACT}_{\text{NCD}}$ is capable of simulating lag times between exposure and disease. All the statistical models that have been described in the previous paragraphs that are used to simulate individual histories for risk factor exposures include year and age as predictors. Therefore, for any synthetic individual past exposures to risk factors are available during the simulation and are used for the calculation of synthetic individuals' disease risk (please refer to section 2.4 on the next page).

Although this approach allows for different lag times to be modelled for any exposure and disease combination, given the lack of evidence for some of them, I followed a simpler approach. I assumed a mean lag time of five years between any CVD related exposure and CVD. I also assumed a mean lag time of eight years between any cancer related exposure and cancers except for smoking. The effects of smoking on lung and gastric cancer appear to be cumulative and the risk gradually declines after cessation (section 1.3.1 on page 18). I modelled this gradual decline and I considered a 5-year mean lag time; hence the risk starts its decline five years after successful smoking cessation. Given the uncertainty of these assumptions I allow them to vary stochastically following a binomial distribution (please refer to section 2.6 on page 62 and table A.1 on page 187).

The 5- and 8-year mean lag times were also selected for consistency with mean observation times of cohorts and randomised control trials that were used to extract the risk of each exposure. Cohort studies and randomised control trials only seldom report relative risks as a function of time. Therefore, the reported relative risks correspond to the mean observation time of each study. For instance, consider a risk factor A with 5-year lag time for full reversibility. The excess risk will gradually decline in the 5-year period after exposure to A stops. Now consider a randomised control trial that trialled an intervention against risk factor A and lasted five years. The trial participants would be under observation usually for a different duration because of different recruitment dates. Therefore, not all of them will experience the full risk reduction and the reported results from the trial will be accurate only for the mean period of observation of its participants. Hence, it is important for the lag times in $\text{IMPACT}_{\text{NCD}}$ to be similar to the mean observation times of the studies from which the relative risks were extracted. Although lag times have been considered for some comparative risk assessment[170, 171], to my knowledge $\text{IMPACT}_{\text{NCD}}$ is the only simulation model that models different lag times for CVD and cancers, in accordance with existing evidence (section 1.3 on page 17).

2.3.4   *Birth engine*

The birth engine represents step 4 in figure 2.1 on page 45. The ONS principal assumption fertility projections for England are used to estimate the number of new synthetic individuals entering the model through birth, in every simulated year.[245] The newborn synthetic individuals have no behavioural risk factors and inherit the percentile ranks for the biological risk factors from a random member of the household, in which the newborns are born. The birth engine only becomes important for simulations with a simulation horizon longer than 30 years.

2.4   DISEASE MODULE

The disease module contains the last three steps of the model (figure 2.1 on page 45). For each synthetic individual aged 30 - 84, the risk (probability) to develop each of the modelled diseases is estimated in step 5 conditional on the exposure to relevant risk factors. The step ends by selecting synthetic individuals to develop the modelled diseases. Finally in steps 6 and 7, the risk of dying from one of the modelled diseases or any other cause is estimated and applied. Steps 2 to 7 are then repeated for the surviving individuals until the simulation horizon is reached.

2.4.1   *Estimating the annual individualised disease risk*

Step 5 in figure 2.1 on page 45 is the step where annual disease incidence is estimated. First, the exposure levels of all the modelled risk factors are transformed to relative risks for the modelled diseases using information from published studies (table 2.2 on page 70). This is performed for all synthetic individuals with ages between 30 and 84, and the lag times define the simulation year from which the exposure levels are extracted. Then, IMPACT$_{NCD}$ estimates what would be the probability of the synthetic individuals developing the disease conditional on their age and sex, if their exposures were at optimal levels. Finally, it uses this probability to calculate the individualised annual probability of a synthetic individual developing a specific disease conditional on his/her relevant past risk exposures. This is a three stage process.

1. The proportion of disease incidence attributable to the modelled risk factors is estimated by age and sex, assuming a lag time.

2. Then, the estimated proportion from the previous step is subtracted from the total disease incidence, assuming multiplicative risks.

3. The probability of developing the disease is estimated for each synthetic individual. Then, it is used in an independent Bernoulli trial to identify those synthetic individuals, who will develop the disease in the specific simulated year.

In the following paragraphs I will describe in more detail these three stages, which are performed separately for each of the modelled diseases.

STAGE 1:   This stage is based on the PAF, an epidemiological method that measures the proportion of the disease burden that can be attributed to a specific risk factor exposure in a population.[246] In other words what proportion of the disease burden could have been eliminated from the population if nobody was exposed to the risk factor. Its formula is:

$$PAF = \frac{p_e(RR - 1)}{p_e(RR - 1) + 1} \qquad (2.1)$$

Where:
- p_e the prevalence of the risk factor in the population
- RR the associated relative risk of the risk factor for the disease that the PAF is calculated for

In a microsimulation setting for a disease with k associated risk factors and assuming multiplicative risk factors formula (2.1) can be written as:

$$PAF = 1 - \frac{n}{\sum_{i=1}^{n}(RR_1 * RR_2 * \cdots * RR_k)} \qquad (2.2)$$

Where:
- i the synthetic individuals
- n the number of synthetic individuals
- $RR_1 * RR_2 * \cdots * RR_k$ the relative risks of the risk factors associated with the disease

IMPACT$_{NCD}$ uses formula (2.2) to estimate the proportion of the disease incidence attributable to the modelled risk factors, by age and sex. I defined the optimal levels for each exposure below which no excess risk exists, from the same studies that I extracted the relative risks.

STAGE 2:   In this stage IMPACT$_{NCD}$ estimates the disease incidence not attributable to the modelled risk factors using the formula:

$$I_{theoreticalminimum} = I_{observed} * (1 - PAF) \qquad (2.3)$$

Where:
- $I_{observed}$ the disease incidence
- PAF the PAF estimated from formula (2.2) in stage 1
- $I_{theoreticalminimum}$ the disease incidence if all the modelled risk factors were at optimal levels

The theoretical minimum incidence is calculated by age and sex only in the initial year of the simulation and it is assumed constant thereafter. Therefore, I assume that the only drivers of time trends in age-standardised disease incidence are the modelled risk factors.

STAGE 3:   Assuming that $I_{theoreticalminimum}$ is the baseline annual probability of a synthetic individual to develop the disease for his/her age and sex due to non-modelled risk factors (i.e. genetics), the individualised annual probability developing the disease, $\Pr(disease \mid age, sex, exposures)$, conditional on his/her exposure to risk factors is:

$$\Pr(disease \mid age, sex, exposures) = I_{theoreticalminimum} * RR_1 * RR_2 * \cdots * RR_k \qquad (2.4)$$

Formula (2.4) is applied every simulated year, for every non-diseased synthetic individual aged 30 to 84 and for every disease. Then independent Bernoulli trials select the synthetic individuals that will develop the disease in every simulated year.

For the initial year of the simulation, some synthetic individuals need to be allocated as prevalent cases for each of the modelled diseases. IMPACT$_{NCD}$ uses disease prevalence as input and randomly allocates synthetic individuals as prevalent cases in the synthetic population. I assumed that the concentration of risk factors is higher among the prevalent cases compared to the general population. Therefore, prevalent cases are sampled independently from the synthetic individuals in the population with weights proportional to their relevant exposures.

2.4.1.1 *Disease incidence and prevalence*

The previous algorithm requires disease incidence and prevalence for the initial simulation year as inputs. As I have explained in section 1.1 on page 7 and section 1.2 on page 13, with the exception of cancer incidence, these are poorly quantified for England. However, a few sources exist although their risk of bias is high because they are either self-reported, or are indirect estimates from health care data. Therefore, I opted for a modelling solution to synthesise all the available information from the available sources and model incidence and prevalence using the mathematical relation between disease incidence, mortality, and prevalence.

I used the WHO model named DisMod II for this. DisMod II is a multistate life table model that is able to estimate the incidence, prevalence, mortality, fatality, and remission of a disease, when information about at least three of these indicators is available.[247] A similar approach has been followed by the Global Burden of Disease team and others.[248–255]

For CHD, I used as inputs for DisMod II: 1. the ONS reported CHD mortality rates (International Classification of Diseases, version 10 (ICD10): I20-25) for England in 2006 by age group and sex;[256] 2. the self-reported prevalence rates of CHD from HSE2006 by age group and sex;[207] 3. the sum of angina incidence rate estimates from primary care data and AMI incidence rate estimates from mortality and hospital statistics by age group and sex.[7, 92] I assumed CHD as an incurable chronic disease, therefore I set the remission rate to zero. Hence, the DisMod II output incidence rates were actually the incidence rates for the first ever manifestation of angina or AMI excluding any recurrent episodes. DisMod II allows past trends to be included as model inputs for more accurate estimations. I assumed that incidence and case fatality had been declining by 3 % (relative), over the last

20 years, based on findings by Smolina *et al.*[7] The derived CHD incidence, prevalence and case fatality were used as inputs for IMPACT$_{NCD}$. The approach for stroke was very similar to CHD. The Lee *et al.* study was used to extract incidence rates and trends.[13]

For lung and gastric cancers the incidence and survival for the first and fifth years from the cancer registries are reported by the ONS on an annual basis. However, prevalence, remission, and case fatality are not recorded in the cancer registries. Hence, I used DisMod II to estimate them. I only will describe here the process for lung cancer since that for gastric cancer was similar. Specifically, I used: 1. the ONS reported lung cancer mortality rates (ICD10: C34) for England in 2006 by age group and sex;[256] 2. the ONS reported lung cancer incidence rates (ICD10: C34) for England in 2006 by age group and sex;[257] 3. from the ONS reported first and fifth year survival rates,[258] I extrapolated the tenth year survival rates by age group and sex, assuming survival follows a Weibull distribution. Then I assumed the tenth year survival rate equals remission rates. With these as inputs DisMod II estimated lung (and gastric) cancer prevalence and case fatality rates that are used by IMPACT$_{NCD}$.

2.4.2   *Simulating disease histories*

In the previous step, IMPACT$_{NCD}$ estimates disease incidence. In the current one (step 6 in figure 2.1 on page 45), IMPACT$_{NCD}$ simulates important aspects of the modelled diseases. Since in the current phase of development IMPACT$_{NCD}$ is used to model primary prevention policies, a detailed disease history module is not necessary. Therefore, only the first ever episodes of CHD and stroke are modelled explicitly and not recurrent episodes.[30] For the moment, in this step IMPACT$_{NCD}$ models the observable spike of short term (30 days) mortality after the first event of AMI or stroke. The 'Coronary heart disease statistics 2012 edition' report was used to extract data regarding short term mortality.[92] Furthermore, for cancers, remission is modelled using DisMod II output remission rates by age and sex (section 2.4.1.1 on page 58).

2.4.3   *Simulating mortality*

So far, I have described how the synthetic population is developed and evolved over time (steps 1 to 4 in figure 2.1 on page 45), then how the risk of developing disease is estimated and disease progress (steps 5 to 6 figure 2.1 on page 45). Finally, the model simulates the resulting mortality in the last step of the algorithm (step 7 in figure 2.1 on page 45). All synthetic individuals are exposed to the risk of dying from any of their acquired modelled diseases or any other non-modelled cause in a competing risk framework. The algorithm behaves differently depending on the age and life course trajectory of the synthetic individual.

---

30 This step may be expanded in the future to explicitly model health care interventions.

For synthetic individuals with ages 0 to 29 or 85 to 99,[31] IMPACT$_{NCD}$ uses all-cause mortality rates by age, sex, and QIMD to inform independent Bernoulli trials and select synthetic individuals that die in every simulated year. For years 2006 to 2013 in the simulation, I used the observed mortality rates as they were reported by the ONS.[256] For years after 2013 observed mortality was not available. Therefore, I fitted functional demographic models[32] to the ONS reported annual mortality rates from years 2002 to 2013, stratified by sex and QIMD. I used these models to forecast mortality rates up to the simulation horizon using the R package 'demography'.[260]

2.4.3.1   *Disease-specific mortality*

For synthetic individuals with ages between 30 and 84 all-cause mortality was decomposed into modelled diseases specific mortality and any other-cause mortality. The former applies only to the prevalent cases of each modelled disease in the synthetic population. For this, disease-specific case fatality rates by age and sex that were estimated by DisMod II (section 2.4.1.1 on page 58) are used for further calculations. Specifically, I assumed that CHD, stroke, lung and gastric cancers case fatalities are improving by 4 %, 4.5 %, 3 %, and 2.5 % (relative) respectively, every year. Furthermore, I assumed that there is a constant socioeconomic gradient in case fatalities by QIMD level of approximately 5 %, 3 %, 3 %, 4 % for CHD, stroke, lung and gastric cancers respectively, that is halved for ages over 70. The socioeconomic gradient forces the more deprived to experience worse disease outcomes. These assumptions are based on extensive empirical evidence.[21, 92, 107–110, 261] After these calculations, the modified case fatality rates are used in independent Bernoulli trials to select prevalent cases that die from the modelled disease in every simulation year.

2.4.3.2   *Other-cause mortality*

For mortality from any other cause, a process similar to the one described for ages 0 to 29 and 85 to 99 is followed (section 2.4.3 on page 59). However, this time modelled disease-specific mortality (i. e. CHD, stroke, lung, and gastric cancers mortality) is removed from the observed mortality before forecasting to avoid double counting.

2.4.3.3   *Effect of smoking and diabetes mellitus on other-cause mortality*

The DECODE study suggested that diabetes mellitus increases the risk of non-CVD mortality.[262] Similarly, the 'male British doctors' study suggested that smoking increases the risk of non-CVD and non-lung cancer mortality.[263] IMPACT$_{NCD}$ models these findings by inflating the mortality rate for smokers and diabetics based on findings from these two studies. This is performed again using a PAF approach and formulas equation (2.2) on page 57 and equation (2.3) on page 57.

---

31 IMPACT$_{NCD}$ has a maximum age limit of 100 years.

32 Functional demographic models are generalisations of the Lee Carter demographic model, influenced by ideas from functional data analysis and non-parametric smoothing.[259]

2.4.3.4   *Randomising the order mortality is calculated*

The sequence that mortality from each of the modelled diseases or any other cause is calculated is important and may introduce bias. For example, consider a synthetic individual that is a prevalent case of CHD and lung cancer simultaneously. If in each simulated year, CHD mortality was always calculated before lung cancer mortality, this would decrease the probability of the individual to die of lung cancer. This is a well known limitation of discrete time microsimulations.[196] IMPACT$_{NCD}$ randomises the order that mortality from each cause is calculated to reduce this bias.

2.4.4   *Algorithm repeat*

Finally, synthetic individuals who remain alive after this step progress to the next year and start again from step 1 (figure 2.1 on page 45), unless the simulation horizon has been reached. IMPACT$_{NCD}$ records incidence, prevalence, and mortality for the modelled diseases for every simulated year.

2.5   SCENARIO SPECIFICATION

The method I described so far, is used to model the baseline scenario. Primary prevention policies can then be modelled as counterfactual scenarios, by altering the baseline scenario. I will discuss the justification of specific scenarios and the parameters that define them in the relevant result chapters. I will present here an overview of the available options within IMPACT$_{NCD}$ framework. In general, primary prevention policies can be modelled through their effects on the relevant risk factor exposures, in three different ways:

OPTION 1:   Population-wide interventions can be modelled by altering the intercept or the coefficients of the regression equations that are used to estimate life histories (section 2.3.2.2 on page 50). For example, when continuous risk factors are considered, adding or subtracting from the intercept increases or decreases the related risk factor for each synthetic individual; therefore, the mean of the risk factor for the whole population. Altering the year coefficient accelerates, decelerates or reverses the trend for the whole population. Likewise, altering the QIMD coefficients or/and the coefficient of the interaction between year and QIMD can simulate differential effects and trends by QIMD. A similar approach sometimes is possible also for the non-continuous risk factors. The benefit is that by just altering a few parameters the changes are translated down to individual level characteristics in a computationally efficient way.

OPTION 2:   Targeted interventions can be modelled by selecting synthetic individuals with a specific trait or combination of traits, and apply an intervention to them. For example, to simulate the effect of statins a simple approach would be to randomly select 30 % of the synthetic individuals with total cholesterol higher than 4 mmol/l not currently on

statins; and apply a 25 % reduction of their total cholesterol between steps 4 and 5 (figure 2.1 on page 45).

OPTION 3:   Some hybrid combination of the previous methods or some more complex approaches have the time slow down, stop in a specific year, or running backwards to simulate 'disaster' scenarios.

2.6   HANDLING UNCERTAINTY IN THE MODEL

There are four sources of uncertainty in decision modelling[264–266]:

STOCHASTIC UNCERTAINTY: This is the random variability in outcomes for identical synthetic individuals. For example, consider two identical synthetic individuals both having a probability of 5 % to develop CHD at a certain year in the simulation. Two independent Bernoulli trials are performed to decide which of them will develop CHD, or not. The outcome of the Bernoulli trials is random. This type of uncertainty is also known as Monte Carlo error, or first order uncertainty.

PARAMETER UNCERTAINTY: This is the uncertainty of the parameters that are used as modelled inputs. For example, consider smoking relative risk for CHD. The true relative risk is unknown and depends on a plethora of known and unknown factors. Epidemiological studies can only approximate the true relative risk. Furthermore, epidemiological studies quantify only the sampling error of their estimates, leaving other sources of bias in the qualitative realm. Parameter uncertainty is sometime called second order uncertainty.

HETEROGENEITY: This is the variability that can be explained by individual characteristics. For example, smoking relative risk for CHD may differ by age of the smoker, duration, and intensity of smoking. Heterogeneity is embedded in microsimulations, but requires substantial effort to be modelled in a macrosimulation setting.

STRUCTURAL UNCERTAINTY: This uncertainty arises from model structure and modelling decisions and assumptions. For example, in discrete time microsimulations the sequence of the events is usually preprogrammed in each cycle. This simplifies model building, however, in the real world the sequence of events is seldom fixed.

2.6.1   *An illustrative example*

The following example illustrates the different types of uncertainty. Consider a simple microsimulation with only one input, a 5 % annual risk for CHD. If this risk is applied to all synthetic individuals and randomly draws from a Bernoulli distribution with $P = 5\,\%$ to select those who will manifest the disease, only stochastic uncertainty is considered. If the annual risk for CHD is further parametrised to be conditional on individual characteristics (i. e. age, sex, exposure to risk factors), then individual heterogeneity is considered. Going

one step further, if the uncertainty of the annual CHD risk is incorporated in the microsimulation because for example it is informed by a cohort study with a known sampling error, then parameter uncertainty is considered. Finally, structural uncertainty may stem from the fact that the risk factors that were modelled may not be truly associated with CHD. From the four types of uncertainty, only parameter uncertainty and possibly structural uncertainty may be reduced from better future research.

2.6.2 *Quantifying uncertainty*

Quantifying overall model uncertainty from all four types of uncertainty is not an easy task and only rarely do models incorporate a full uncertainty analysis in their estimates. Some argue that full uncertainty analysis is sometimes pointless especially when the policy maker has no power to reduce the uncertainty by providing for example more accurate inputs.[266] Nevertheless, acknowledging information gaps is a huge advantage of modelling because it can guide future research and help policy makers to make informed decisions given the level of certainty in the model.[156, 266]

IMPACT$_{NCD}$ implements a second order Monte Carlo approach to quantify uncertainty in the model for each scenario. This approach allows stochastic uncertainty, heterogeneity, and parameter uncertainty to be propagated in the model outputs. Second order Monte Carlo is a method to study uncertainty by repeating each scenario simulation multiple times; for each iteration a different set of input parameters is used. Essentially, every microsimulation is a first order microsimulation because the outcome is assessed for each synthetic individual separately and each synthetic individual has a different set of traits. Therefore, it combines stochastic uncertainty and heterogeneity. In a second order Monte Carlo a set of input parameters is sampled from their respective uncertainty distributions and is then used in a first order Monte Carlo.[264, 265]

IMPACT$_{NCD}$ repeats each modelled scenario 1000 times. A different set of input parameters is sampled from uncertainty distributions in each iteration. I assumed log-normal distributions for relative risks and hazard ratios, normal distributions for parameters informed by coefficients of linear regression equations, and PERT distributions[33] for inputs based on user assumptions. The choice of these distributions is common practice in modelling.[268] For relative risks and hazard ratios, the distributions were bounded above one when their mean was above one and vice versa. For example, I did not allow high blood pressure to reduce the risk for CVD under no circumstances. All parameter input distributions that are used in IMPACT$_{NCD}$ are summarised in table A.1 on page 187. Because of this approach, IMPACT$_{NCD}$ returns a distribution for each model output instead of a unique value. I summarise the output distributions by reporting medians for point estimates and interquartile ranges (IQRs) (in the form of first and third quartiles) for the estimated uncertainty. The structural uncertainty is not quantified in the model outputs. However, IMPACT$_{NCD}$ is grounded on fundamental epidemiological ideas and well estab-

---

33 The PERT distribution is a version of the Beta distribution. It is very useful in modelling experts' estimates because it can be parametrised with a minimum, a maximum, and a 'most likely' (mode) value. A forth parameter can used to reflect experts' 'level of belief' to their estimates.[267]

lished causal pathways; therefore, I consider this type of uncertainty relatively small in this particular case.

Two modelling decisions are particularly important for understanding uncertainty in IMPACT$_{NCD}$. The first is that for parameter uncertainty, only the sampling error from the epidemiological studies was considered. This decision was made because epidemiological studies only quantify uncertainty from sampling error and an attempt to quantify i. e. bias would be highly subjective. The second decision is more subtle. Many scenarios share the vast majority of input parameters and differ only in a few of them. For example consider two tobacco control policies, one reducing smoking prevalence by 5 % and the other by 10 %. Modelling these two policies would require all other model inputs to be the same, except for smoking prevalence. However, all shared input parameters have an estimated uncertainty which may be of importance. If the interest is in a relative comparison between the two policies, then the uncertainty from the share parameters is not important because it mostly cancels out. However, if the interest is in estimating absolute values of prevented cases of CVD for example, then the uncertainty from the shared parameters becomes important. IMPACT$_{NCD}$ uses the same input parameter values for the shared input parameters and the same initial population in each iteration. This allows for 'paired' comparisons between scenarios that result in substantially less estimated uncertainty than the uncertainty of isolated scenarios. It is worth noting that despite the initial population being the same for each scenario in each iteration, life courses are not. The same synthetic individual may have a different life course under each modelled scenario, not only as a result of the modelled policy but also because of chance.

2.7   MODEL OUTPUTS

2.7.1   *Policy effectiveness metrics*

IMPACT$_{NCD}$ tracks incidence, prevalence, and mortality of the modelled diseases by year, and it can stratify them by any of the modelled individual traits. When comparing effectiveness of policies though, the derived metrics of cases prevented or postponed and deaths prevented or postponed appear more useful because they summarise the estimated effectiveness of the policy in two numbers. Cases prevented or postponed from a policy are measured by subtracting the number of incident cases observed under the policy scenario, from the number of incident cases observed under the baseline scenario, over a period of time. The calculation is similar for the deaths prevented or postponed.

2.7.2   *Policy equity metrics*

Please refer to section 1.4.4 on page 29 for a gentle introduction on measuring socioeconomic inequalities in health. In the next paragraphs I will reverse the perspective from population inequalities to policy equity and I will describe the policy equity metrics that IMPACT$_{NCD}$ estimates.

2.7.2.1 *Absolute and relative equity slope indices*

Cases and deaths prevented or postponed can be stratified by QIMD. However, because QIMD has five levels, comparison of the equity of a policy is not straightforward. Mackenbach *et al.* developed two more sophisticated measures of socioeconomic inequalities in health, the slope index of inequality (SII) to measure absolute inequality and the relative index of inequality (RII) to measure relative inequalities. Both are regression based measures that take into account the size and relative socioeconomic position of groups.[115]

In the spirit of slope and relative index of inequality I developed the absolute equity slope index and the relative equity slope index; two regression based metrics, to measure the impact of the modelled interventions on absolute and relative socioeconomic health inequalities. However, instead of directly measuring inequalities in a population in the way that SII and RII do, they measure the impact of an intervention on existing inequalities.

The basic principles of the metrics are illustrated in this simplified example. Consider the simple example of a population that consists of only two mutually exclusive and same sized socioeconomic groups, the 'deprived' and the 'affluent'. The two groups experience different incidence of a disease; supposedly, 50 and 10 incident cases among the deprived and the affluent, respectively, every year. Hence, the absolute socioeconomic inequality for disease incidence is 50 − 10 = 40 cases and the relative socioeconomic inequality is 50/10 = 5. If a hypothetical intervention 'A' prevents the same number of cases in both groups, absolute inequality will remain stable. If intervention 'A' prevents more cases in the affluent group, absolute inequality will increase and vice versa. For relative inequality to remain stable, the decrease in cases need to be proportional to the observed number of cases. For example, a hypothetical intervention 'B' that reduces 10 % of cases in each group will have no effect on relative inequality. If the proportional reduction is higher in the affluent group compared to the deprived, then relative inequality will increase and vice versa. The same principles apply when the population is split into five unequally populated socioeconomic groups as with QIMD. However, in this case, the absolute and relative inequalities cannot be defined by a simple subtraction or deviation and the SII and RII have to be used. If an intervention prevents an equal number of cases for every QIMD, SII will remain unchanged. If the proportional reductions of cases for every QIMD are equal, RII will remain unchanged.[34]

Inspired by the SII and RII the absolute equity slope index is the slope of the regression line fitted in the number of cases prevented or postponed by an intervention, by QIMD (dependent variable). To account for the different population sizes in each QIMD each group is given a score, called ridit score, which reflects the average cumulative frequency of the group (independent variable).[269] A positive slope means that the intervention prevents more cases in the more deprived QIMD and reduces absolute inequality in the population and vice versa. The magnitude of the slope is proportional to the reduction in absolute inequality. The relative equity slope index is constructed and interpreted similarly, except that the proportion of cases prevented or postponed over the total cases in each

---

34 Assuming that the deaths prevented by the intervention do not change the relative size of the socioeconomic groups.

socioeconomic group is the independent variable, and it measures the effect on relative socioeconomic health inequality.

2.7.2.2 *Equity summary chart*

The equity summary chart presents, in a simple two dimensional chart, the impact of the interventions on disease incidence, and absolute and relative socioeconomic inequality. The horizontal axis represents the number of cases prevented or postponed and the vertical axis represents the decrease (or increase) in absolute inequality. An 'equity' curve divides the graph in two parts. Interventions above the equity curve decrease relative inequality and interventions below it increase relative inequality (please refer to figure 5.2 on page 130 for a plotted example). The equity summary chart and the equity curve have an underlying assumption. For a given overall reduction of disease burden in the whole population attributable to an intervention, there is one and only one way to distribute this reduction among the socioeconomic groups, that can reduce absolute socioeconomic inequality and have no impact on relative socioeconomic inequality. The advantage of the equity summary chart is that presents on a two axes chart the impact of the intervention on disease incidence, absolute, and relative inequalities in agreement with recommendations by health inequalities experts (section 1.4.4 on page 29).

2.8 TECHNICAL SPECIFICATION

IMPACT$_{NCD}$ has been developed in R and is currently deployed in an 80 core server running Scientific Linux v6.2.[209] IMPACT$_{NCD}$ is built around the R package 'data.table', which imports a new heavily optimised data structure in R.[270] Most functions that operate in a data table have been coded in C to improve performance. Each iteration for each scenario is running independently in one of the cores and the R package 'foreach' is responsible for the distribution of the jobs and collection of the results.[271] To ensure statistical independence of the pseudorandom number generators running in parallel, the R package 'doRNG' was used to produce independent random steams of numbers, generated by L'Ecuyer's combined multiple recursive generator.[272, 273]

2.9 DISCUSSION

In this chapter, I have presented the IMPACT$_{NCD}$ framework in an accessible way for non-technical readers, while I provide access to the source code of the model for those familiar with R programming language. The Brighton declaration recognised communicating model aspects and model uncertainty to policy makers as a major challenge in the field.[161] I believe that my approach of presenting IMPACT$_{NCD}$ is addressing this challenge to a certain extent.

In section 1.7 on page 40 I stated that one of the aims of my thesis was to build a simulation model that would be reusable, transparent, and comprehensive. IMPACT$_{NCD}$ is reusable because it simulates a virtual world that is informed by the current knowledge

about the real world. Any primary prevention policies, additional NCDs or even health care can be modelled as extensions to the current model which speeds up development time. IMPACT$_{NCD}$ is transparent because it is open-source. Anyone can see the source code, use it, or improve it without prior permission. It is also transparent because all of its algorithms are presented as a narrative, which does not require advance technical knowledge to be understood. Finally, IMPACT$_{NCD}$ is a comprehensive model because it is grounded on well accepted principles of traditional and social epidemiology to model the interplay of the wider determinants of health with risk factor exposures and disease burden.

Unlike other microsimulations, data requirements for IMPACT$_{NCD}$ are limited to a health survey of the population and population vital statistics. All other input sources are published epidemiological studies. The studies that inform IMPACT$_{NCD}$ were selected because of their high quality. When available, individual level meta-analyses were preferred over dose response meta-analyses of summary statistics, and dose response meta-analyses were preferred over meta-analyses of summary statistics and primary epidemiological studies. All IMPACT$_{NCD}$ sources have been summarised in table 2.2 on page 70. In some cases, more than one suitable meta-analyses were identified to inform IMPACT$_{NCD}$ sources (i. e. to inform the relative risk of environmental tobacco smoking for lung cancer, or the relative risk of fruit and vegetable consumption for lung cancer). I report this in the column named 'Comments' in the aforementioned table. In all such cases, the differences in the parameter estimates where very small, especially compared to the overall uncertainty of the model. It was not possible to identify any suitable meta-analyses to inform the relative risk of non-CVD mortality for diabetics and non-CVD non-lung cancer mortality for smokers. In those cases I identified two cohort studies that reported these relative risks.[262, 263] Finally, I rejected meta-analyses that did not report their results by age group, sex, and specific disease (i. e. separately for CHD and stroke instead of overall CVD) when alternative meta-analyses existed that reported their results stratified by age group, sex, and specific diseases.

The core of the disease module in IMPACT$_{NCD}$ is based on the PAF. PAF is a very useful epidemiological idea and is used in every comparative risk assessment and many simulation models, like SimSmoke, IMPACT and the DYNAMO-HIA (section 1.6.1 on page 35). However, the equation (2.1) is valid only when there is no confounding in the exposure/disease association,[274] an assumption that rarely holds for NCDs. Many alternatives have been proposed to relax this assumption, however none of them have been widely used due to their complexity and data requirements.[274–276] Seminal comparative risk assessments have used slightly modified versions of equation (2.1) assuming multiplicative risks and using relative risks adjusted for confounding.[30, 32, 171] Rückinger *et al.* compared the performance of the original PAF formula with some of the proposed alternatives and showed that indeed the original formula performs badly. Unfortunately, they did not include in the comparison a version of the original formula with the multiplicative risk assumption.[277] If they had, it would have performed as well as the most accurate and advanced method in the estimation of the total burden of disease attributable to all the

risk factors that have been included in the study.[35] In IMPACT$_{\text{NCD}}$ I also assumed multiplicative risk factors and I used adjusted relative risks to estimate the theoretical minimum disease incidence if all exposures were at optimal level (section 2.4.1 on page 57). The recalculation of Rückinger *et al.* results provide some evidence that my approach does not introduce substantial bias.

The main IMPACT$_{\text{NCD}}$ assumptions and limitations have been summarised in table 2.1 on the facing page and they are further discussed in section 8.5 on page 173. Other assumptions had to be made to address smaller gaps in knowledge and have been mentioned in their relevant paragraphs in this chapter. They are further discussed in the relevant result chapters. In the next chapter I will examine IMPACT$_{\text{NCD}}$ validity and I will explore the impact of these assumptions on model outputs.

---

[35] From table 3 in Rückinger *et al.* [277] the best performing methods have a sum of 90.3 (columns 4-6). The original formula has a sum of 259.8 (column 1) and the original formula with adjusted odd ratios a sum of 194 (column 2). A recalculation of the sum of the second column assuming multiplicative risks gives $sum = 1 - (1 - 0.588) * (1 - 0.228) * (1 - 0.336) * (1 - 0.23) * (1 - 0.173) * (1 - 0.263) * (1 - 0.122) = 91.3$, similar to the best performing methods.

Table 2.1: IMPACT$_{NCD}$ key assumptions and limitations.

Population module	Migration flows are not considered
	Social mobility is not considered
	QIMD is a marker of relative area deprivation with several versions since 2003. I consider all versions of QIMD approximately unchanged
	I assume all salt that is consumed is excreted from urine and all urine sodium origins from salt consumption
	I assume that the surveys used, are truly representative of the population
Disease module	I assume multiplicative risk effects
	I assume log-linear exposure – response relationship for the continuous risk factors
	For CVD, I used modelled incidence rates derived from mortality and prevalence data
	I define CVD as the sum of CHD and stroke cases
	I assume that the effects of the risk factors on incidence and mortality are equal and risk factors are not modifying survival
	I assume 5-year mean lag time for CVD and 8-year for lung and gastric cancers (except for the cumulative effect of smoking on lung and gastric cancers in which case lag time is assumed five years)
	I assume 100 % risk reversibility
	I assume that trends in disease incidence are attributable only to trends of the relevant modelled risk factors
	Only well accepted associations between behavioural and biological risk factors that have been observed in longitudinal studies are considered. However, the magnitudes of the associations were extracted from a series of nationally representative cross-sectional surveys (HSE)
	For lung and gastric cancers, I assume that survival 10 years after diagnosis equals remission

Abbreviations: cardiovascular disease (CVD); coronary heart disease (CHD); Health Survey for England (HSE); quintile groups of Index of Multiple Deprivation (QIMD).

Table 2.2: IMPACT$_{NCD}$ main data sources.

Parameter	Source	Comments
Initial population structure	Office for National Statistics. Published ad hoc data: health, requests during December 2014. 2014.[256]	Mid-year population estimates for England. Stratified by age, sex, QIMD for years 2002-2013
Fertility rates	Office for National Statistics. National population projections, 2012. 2013.[245]	Principal assumption fertility projections for England. Stratified by age for ages 15 to 46
Mortality rates	Office for National Statistics. Published ad hoc data: health, requests during December 2014. 2014.[256]	Recorded mortality in England for years 2002-2013. Stratified by age, sex, QIMD, and cause of death
Synthetic individuals' exposure to risk factors through their life course	HSE2001-2012 datasets[204–208, 232–238]	Anonymised, individual level datasets from an annually repeated cross-sectional health survey in England. Years 2001-2012
Salt consumption relative risk for gastric cancer incidence	Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington, DC: World Cancer Research Fund and American Institute for Cancer Research, 2007 . Figure 4.6.1	Meta-analysis of 2 cohort studies. Both studies adjusted for age, sex, and smoking. One was also adjusted for non-green/yellow vegetable intake and the other for education, gastric disorders, and history of gastric cancer in the family. None was adjusted for Helicobacter pylori
Effect of salt consumption on SBP	Mozaffarian D et al. Global sodium consumption and death from cardiovascular causes. N Engl J Med 2014;371:624–634. Text S1 in the appendix	Meta-analysis and meta-regression of 103 trials with duration longer than seven days
Ideal salt consumption below which no risk was considered	Mozaffarian D et al. Global sodium consumption and death from cardiovascular causes. N Engl J Med 2014;371:624–634. Text S4 and Table S3 both in the appendix	Evidence from ecological studies, randomised trials and meta-analyses of prospective cohort studies. Intake levels associated with lowest risk ranged from 1.5 g/d to 6.0 g/d

...continued

Parameter	Source	Comments
Active smoking relative risk for CHD and stroke	Ezzati M et al. Role of smoking in global and regional cardiovascular mortality. Circulation 2005;112:489–497. Table 1 Model B	Reanalysis of the CPSII prospective cohort study, with 6-year of follow-up. Stratified by age and sex. Adjusted for age, race, education, marital status, 'blue collar' employment in most recent or current job, weekly consumption of vegetables and citrus fruit, vitamin (A, C, and E) use, alcohol use, aspirin use, body mass index, exercise, dietary fat consumption, hypertension, and diabetes at baseline
Active smoking relative risk for lung cancer	Tammemägi MC et al. Selection criteria for lung-cancer screening. N Engl J Med 2013;368:728–736. Table 2	PLCO Cancer Screening Trial. The statistical model estimates the probability of a diagnosis of lung cancer within a 6-year period taking into account smoking intensity and duration
Active smoking relative risk for gastric cancer	González CA et al. Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). Int J Cancer 2003;107:629–634. Table III	EPIC prospective cohort study with 5-year follow-up. Stratified by country. Adjusted for sex, consumption of vegetables, fresh fruits, processed meat, alcohol, body mass index, and educational level
Active smoking relative risk for other-cause mortality	Doll R et al. Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 2004;328:1519. Table 1	Male British doctors prospective cohort study with 50 years follow-up. Age-standardised. $IMPACT_{NCD}$ excludes the excess mortality risk from CHD, stroke, or lung cancer if any of these diseases are explicitly modelled in the simulation
Ex-smoking relative risk for for CHD	Huxley RR et al. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. Lancet 2011;378:1297–1305. Webfigure 8	Meta-analysis of pooled estimates from 19 prospective studies. Multiply adjusted

...continued

Parameter	Source	Comments
Ex-smoking relative risk for stroke	Wolf PA et al. Cigarette smoking as a risk factor for stroke: the Framingham study. JAMA 1988;259:1025–1029.	The Framingham study. Prospective cohort study. Stroke risk decreased significantly by two years and was at the level of never-smokers by five years after cessation of cigarette smoking
Ex-smoking relative risk for lung cancer	Tammemägi MC et al. Selection criteria for lung-cancer screening. N Engl J Med 2013;368:728–736. Table 2	PLCO Cancer Screening Trial. The statistical model estimates the probability of a diagnosis of lung cancer within a 6-year period taking into account smoking intensity and duration, and years since smoking cessations
Ex-smoking relative risk for gastric cancer	González CA et al. Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). Int J Cancer 2003;107:629–634. Table IV	EPIC prospective cohort study with 5-year follow-up. Stratified by country. Adjusted for sex, consumption of vegetables, fresh fruits, processed meat, alcohol, body mass index, and educational level
Environmental tobacco smoking relative risk for CHD	He J et al. Passive smoking and the risk of coronary heart disease — A meta-analysis of epidemiologic studies. N Engl J Med 1999;340:920–926. Table 3, adjusted model	Meta-analysis of 10 cohort and case-control studies. Adjusted for important CHD risk factors
Environmental tobacco smoking relative risk for stroke	Oono IP et al. Meta-analysis of the association between secondhand smoke exposure and stroke. J Public Health 2011;33:496–502. Figure 1	Meta-analysis of 20 prospective, case-control, and cross-sectional studies, 13 studies adjusted for important CHD risk factors. The overall effect from all 20 studies was used
Environmental tobacco smoking relative risk for lung cancer	Kim CH et al. Exposure to secondhand tobacco smoke and lung cancer by histological type: a pooled analysis of the International Lung Cancer Consortium (ILCCO). Int J Cancer 2014;135:1918–1930. Table 3	Meta-analysis of individual data from 18 case-control studies. Adjusted for age, sex, race/ethnicity, and study. Study reports odds ratio. I assumed that relative risk is approximately the same. Results are similar to previous meta-analyses by Taylor *et al.* [283]

Parameter	Source	Comments
SBP relative risk for CHD and stroke	Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002;360:1903–1913. Figures 3 and 5	Meta-analysis of individual data from 61 prospective studies. Stratified by age and sex. Adjusted for regression dilution and total blood cholesterol and, where available, lipid fractions, diabetes, weight, alcohol consumption, and smoking at baseline
Total cholesterol relative risk for CHD and stroke	Prospective Studies Collaboration. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55 000 vascular deaths. Lancet 2007;370:1829–1839. Webtable 6 fully adjusted and Figure 3	Meta-analysis of individual data from 61 prospective studies. Stratified by age and sex. Adjusted for regression dilution, age, sex, study, SBP, and smoking
BMI relative risk for CHD and stroke	The Emerging Risk Factors Collaboration. Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. Lancet 2011;377:1085–1095. Table 1 and Figure 2	Meta-analysis of 58 prospective studies. Stratified by age. Adjusted for age, sex, smoking status, SBP, history of diabetes, and total cholesterol
BMI relative risk for gastric cancer	Continuous Update Project report: diet, nutrition, physical activity and stomach cancer. Research rep. World Cancer Research Fund International/American Institute for Cancer Research, 2016 . Table 8, p37	Non-linear dose response meta-analysis for risk of cardia gastric cancer. Adjusted for age, sex, and smoking
Diabetes mellitus relative risk for CHD and stroke	The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010;375:2215–2222. Figure 2	Meta-analysis of 102 prospective studies. Stratified by age. Adjusted for age, smoking status, BMI, and SBP

...continued

Parameter	Source	Comments
Diabetes mellitus relative risk for other-cause mortality	The DECODE Study Group. Is the current definition for diabetes relevant to mortality risk from all causes and cardiovascular and noncardiovascular diseases? Diabetes Care 2003;26:688–696.	DECODE, a collaborative prospective study of 22 cohorts in Europe. Adjusted for BMI, SBP, smoking, and total cholesterol
Physical activity relative risk for CHD and stroke	[69, Tables 10.19 and 10.20]	Meta-analysis of 18 cohort studies for CHD and eight cohort studies for ischaemic stroke. Stratified by age and sex. Adjusted for measurement error, age, sex, smoking, SBP, and total cholesterol
Fruit and vegetable consumption relative risk for CHD	Dauchet L et al. Fruit and vegetable consumption and risk of coronary heart disease: a meta-analysis of cohort studies. J Nutr 2006;136:2588–2593.	Dose response meta-analysis of nine cohort studies. Multiply adjusted. Dauchet *et al.* reported possible publication bias. However, their results are almost identical to meta-analysis from Wang *et al.* (their results were not a affected by publication bias), and to a large cohort study from Oyebode *et al.* [49, 50]
Fruit and vegetable consumption relative risk for stroke	Dauchet L et al. Fruit and vegetable consumption and risk of stroke: a meta-analysis of cohort studies. Neurology 2005;65:1193–1197.	Dose response meta-analysis of seven cohort studies. Multiply-adjusted
Fruit and vegetable consumption relative risk for lung cancer	Vieira AR et al. Fruits, vegetables and lung cancer risk: a systematic review and meta-analysis. Ann Oncol 2016;27:81–96. Figure 2A	Dose response meta-analysis of 14 cohort studies. No extra benefit for consumption higher than 400 g. Evidence of heterogeneity bias and publication or small study bias. Very similar to Wang *et al.* meta-analysis.[53]

Parameter	Source	Comments
Fruit and vegetable consumption relative risk for gastric cancer	Lock K et al. Comparative quantification of health risks. Chapter 9: low fruit and vegetable consumption. Geneva: World Health Organisation, 2004 . Table 9.28	Reanalysis of the Netherlands Cohort study. Stratified by age group. Adjusted for age, sex, smoking, education, gastric disorders, and family history of gastric cancer. I considered a risk only for less than two portions per day.[24]

Abbreviations: body mass index (BMI); coronary heart disease (CHD); Health Survey for England (HSE); quintile groups of Index of Multiple Deprivation (QIMD); Office for National Statistics (ONS); systolic blood pressure (SBP).

3

VALIDATION

3.1 INTRODUCTION

In the previous chapter, I described the inputs, logic, and outputs of IMPACT$_{NCD}$. IMPACT$_{NCD}$ is a complex model that, like any other simulation model, is based on incomplete data, numerous assumptions, and approximations. In this chapter I will present the validation of the model components to explore the impact of the assumptions and approximations on the outputs of the model.

The International Society for Pharmacoeconomics and Outcomes Research jointly convened with the Society for Medical Decision Making Modelling Good Research Practices Task Force has defined validation as "…a set of methods for judging a model's accuracy in making relevant predictions".[287] The Task Force identified five types of validation:

1. Face validity; this consists of experts' evaluation of the model inputs, processes, and outputs.

2. Internal validity; to what extent mathematical equations and coding are correct and consistent.

3. Cross validity; this is a between models comparison of the outputs to assess and explain similarities and differences.

4. External validity; the comparison of model outputs to real world events.

5. Predictive validity; this is similar to external validity except that it requires model outputs to precede the real world event.

Despite the numerous validation types, the Task Force accepted that "it is not possible to specify criteria a model must meet to be declared 'valid' …". However, the Task Force recommended that external and predictive validity are the most desirable and solid validation types. While a model cannot be declared valid, specific model applications can.

In response to the Task Force recommendations Vemer *et al.* insightfully argued that the state of the art validation definition used by the Task Force was inspired by more technical disciplines and lacks flexibility and practicality when applied to health care models. Vemer *et al.* offered an alternative definition; "…the act of evaluating whether a model is a proper and sufficient representation of the system it is intended to represent in view of an application". They further clarified the terms 'proper' as "…in accordance with what is known about the system" and 'sufficient' as "…that the results can serve as a solid basis for decision making". Finally, Vemer *et al.* argued that while a model cannot be declared as 'valid' in general, a practical approach could be to declare a model "valid enough to reliably support a decision to be based on its outcomes".[288]

The aim of this chapter was to assess the validity of all IMPACT$_{\text{NCD}}$ components. I used the International Society for Pharmacoeconomics and Outcomes Research jointly convened with the Society for Medical Decision Making Modelling Good Research Practices Task Force recommendations as a guide, but in the spirit of Vemer *et al.* response.

3.2   METHODS

For the purpose of validation, IMPACT$_{\text{NCD}}$ can be reduced to three distinctive elements: 1. the initial static synthetic population (section 2.3.2.1 on page 46); 2. a process to evolve the synthetic population, and generate life histories (section 2.3.2.2 on page 50); and finally 3. a process to transform changes of risk factor exposures into changes in diseases incidence and mortality (section 2.4 on page 56). In this section I will describe the validation process of all three IMPACT$_{\text{NCD}}$ elements. Face validity is discussed separately in the 'Discussion' section of this chapter (section 3.4 on page 97), as it is more subjective.

The initial static synthetic population was validated by plotting graphs to visually compare risk factor exposure distributions of a random sample of 200 000 synthetic individuals with the corresponding distributions of the survey that was used as the primer. I used mosaic plots for categorical variables and empirical cumulative distribution plots for ordinal and continuous variables. Mosaic plots are graphical representations of a contingency table of two or more categorical variables, using tiles with areas proportional to the frequencies in each cell of the table.[289] I also calculated the linear correlation structure of the survey population and compared it with the linear correlation structure of the synthetic population. Many variables were nominal or ordinal and the continuous variables were not normally distributed. I used Spearman correlations for the continuous variables, polychoric correlations for the categorical variables, and polyserial correlations for comparisons between ordinal and continuous variables.[290, 291] I used the R package 'wCorr' to calculate the linear correlations.[292]

To validate the regression models that were used to evolve the synthetic population over time and simulate life histories of the synthetic individuals, I again used a graphical approach. For continuous risk factors, I plotted the mean exposure of the synthetic population against the mean exposure in the corresponding survey for years 2001 to 2012.[36] Similarly, for nominal and ordinal risk factors I plotted their prevalence in the synthetic population and the corresponding survey.

Next, I assessed how the changes in risk factor exposures over time reflect on disease incidence and mortality trends. Specifically, I first plotted the IMPACT$_{\text{NCD}}$ estimated lung and gastric cancers incidence rates against those reported from the cancer registries for years 2006 to 2013. Then, I plotted the IMPACT$_{\text{NCD}}$ predicted CHD, stroke, lung and gastric cancers mortality rates against those reported from the mortality registry for years 2006 to 2013. Finally, I fitted a Bayesian age-period-cohort model to the data from the mortality registry and I used it to forecast mortality rates up to 2035.[293, 294] Throughout the

---

36 After I removed the boost samples from HSE.

validation process I stratified my analysis by age group, sex, and QIMD whenever relevant and the available data allowed for.

3.3 RESULTS

3.3.1 *Initial synthetic population*

I plotted graphs to compare the exposures of a random sample from the synthetic population to the HSE2006 sample. The graphs serve as internal validation and are presented in appendix B.1 on page 197 to allow unobstructed text flow. It is evident especially in the cumulative distribution graphs and the mosaic plot for diabetes mellitus validation (figure B.12 on page 209) that the method I used to generate the initial synthetic population creates synthetic individuals with a combination of traits not present in the primer survey. Furthermore, the method was able to produce a synthetic population with linear correlation structure similar to the one in the primer survey (figure B.15 on page 212). In general, the graphs suggest that the initial synthetic population is indeed close to reality.

3.3.2 *Risk factor trends (life histories)*

Similarly, for the validation of risk factor trends I compared trends of risk factor exposures in a sample of the synthetic population, to the observed exposure trends in HSE and the Sodium Survey series. I stratified by age group, sex, and when data allowed by QIMD. I present the graphs in appendix B.2 on page 213. Overall, the graphs provide evidence that the regression models that were used for the simulation of individual life histories (section 2.3.2.2 on page 50) have captured trends by age, sex, and QIMD adequately. Therefore, IMPACT$_{NCD}$ captures and simulates the observed dynamics in population exposures.

3.3.3 *Disease incidence trends*

The following graphs validate IMPACT$_{NCD}$ against observed cancer incidence trends. Only the incidence rates for 2006 were used as model input. Therefore, unlike the validation of the population module, this is external validation. IMPACT$_{NCD}$ appears to simulate gastric cancer incidence rate trends accurately. On the contrary, it moderately underestimates the magnitude of the upward trend of lung cancer in women incidence. This may be because IMPACT$_{NCD}$ uses a sex agnostic equation developed by Tammemägi *et al.* to estimate the cumulative risk of smoking.[295] However, smoking may be more harmful in women than men, although the evidence is still inconclusive.[296] The issue and a practical solution is further discussed in section 7.2.3 on page 150. As it was discussed in section 1.1.3.2 on page 8, the true incidence of CVD is largely unknown, so this part of the model cannot be easily validated.
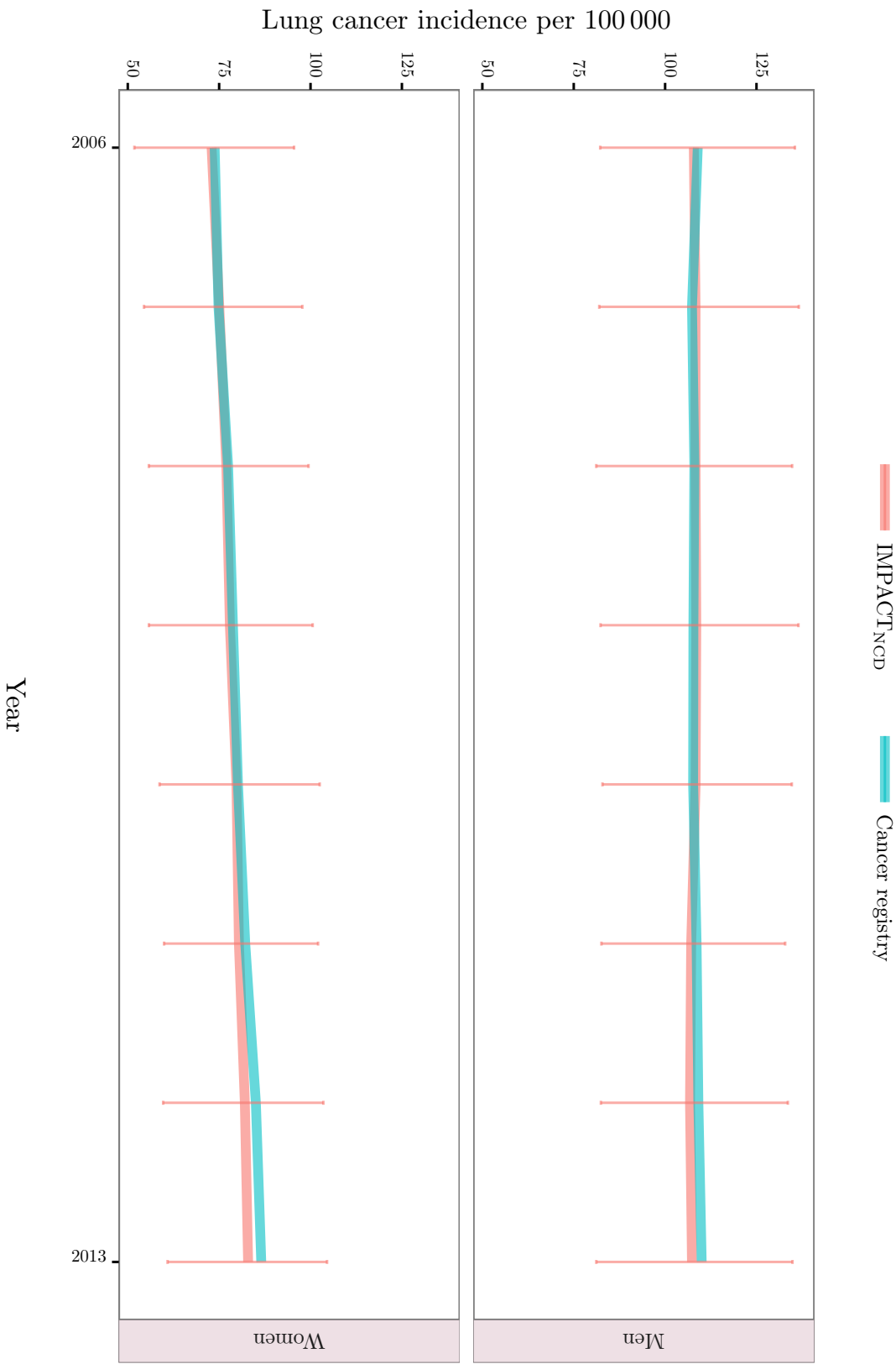
Figure 3.1: Lung cancer incidence rate trends for ages 30 – 84 between years 2006 and 2013. Observed in the population through the cancer registries versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % uncertainty intervals.

Figure 3.2: Lung cancer incidence rate trends for ages 30 − 84 between years 2006 and 2013. Observed in the population through the cancer registries versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % uncertainty intervals.

Figure 3.3: Gastric cancer incidence rate trends for ages 30 – 84 between years 2006 and 2013. Observed in the population through the cancer registries versus IMPACT_NCD synthetic population estimates, stratified by sex. Error bars depict 95 % uncertainty intervals.

Figure 3.4: Gastric cancer incidence rate trends for ages 30 – 84 between years 2006 and 2013. Observed in the population through the cancer registries versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % uncertainty intervals.

3.3.4 *Mortality trends*

Finally, the following graphs conclude the external and cross validation of the disease module. CHD mortality rate trends validation appears to be reasonably good both against the observed mortality trends and the forecasted estimates from the Bayesian age-period-cohort model.

Stroke mortality validation appears to be less accurate than the one for CHD mortality, especially for older ages and comparing to the Bayesian age-period-cohort model forecasts (figure 3.9 on page 89 and figure 3.10 on page 90). This is most likely an artefact. In 2011, ONS updated the software used for cause of death coding, from ICD10 v2001.2 to v2010. One of the consequent changes was that since 2011, vascular dementia deaths have not been coded under cerebrovascular deaths, which led to a reduction in the recorded cerebrovascular deaths in older ages.[297] I was able to adjust for that in the estimates stratified by sex (figure 3.8 on page 88); however, this was not possible for the data stratified by QIMD that were also used to fit the Bayesian age-period-cohort model. The sudden decrease of cerebrovascular deaths from 2011 onwards, influenced the Bayesian age-period-cohort to underestimate the mortality rates forecast for older age groups.

IMPACT$_{NCD}$ appears to overestimate lung cancer deaths. This is more pronounced for older ages and more deprived groups. Given that the lung cancer incidence estimates from IMPACT$_{NCD}$ were rather an underestimation, the overestimation of mortality is probably the result of the case fatality assumptions (section 2.4.3.1 on page 60). On the contrary, gastric cancer mortality trends are reproduced well by IMPACT$_{NCD}$.
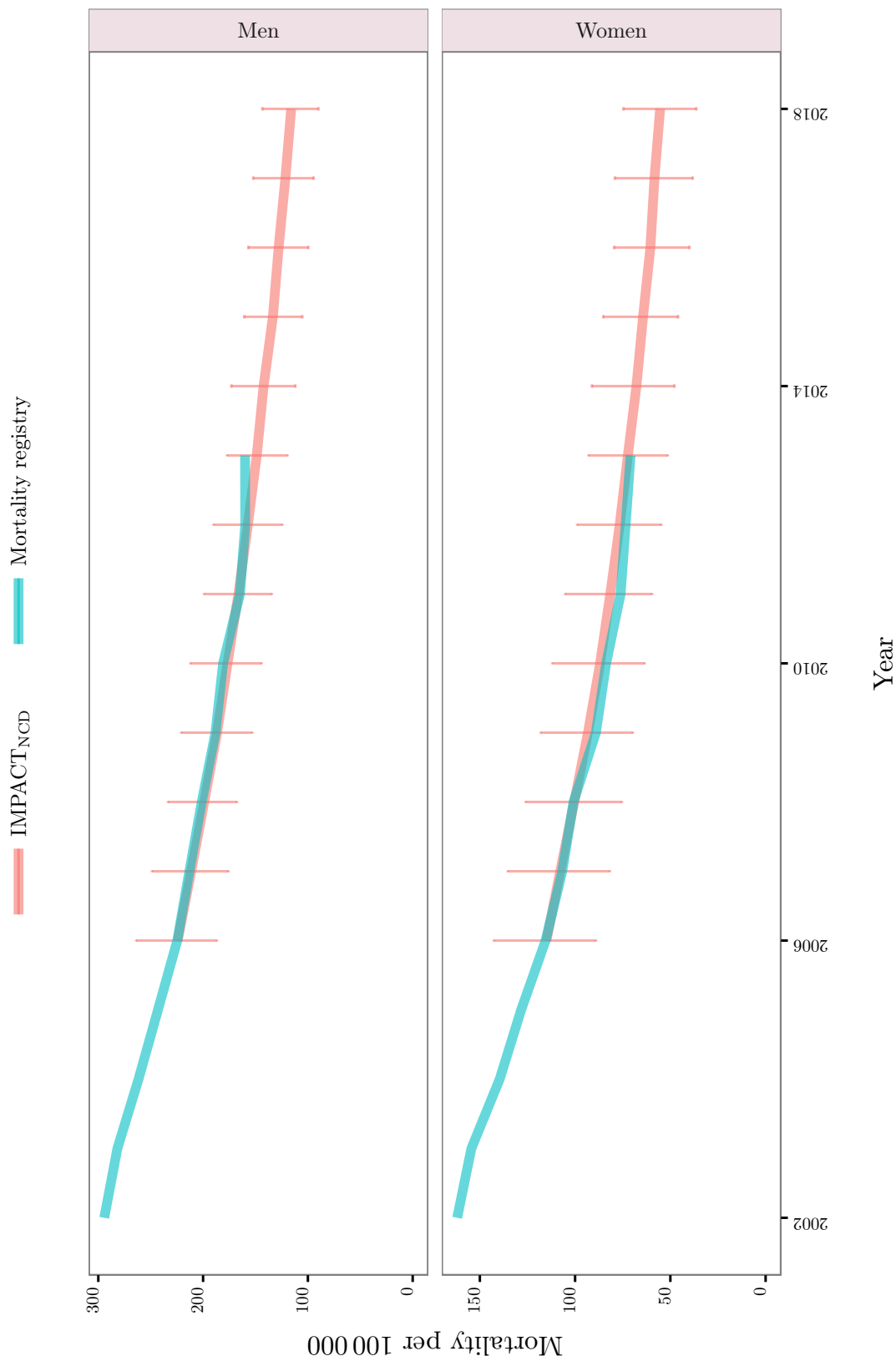
Figure 3.5: Coronary heart disease mortality rate trends for ages 30 – 84 between years 2002 and 2018. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % uncertainty intervals.
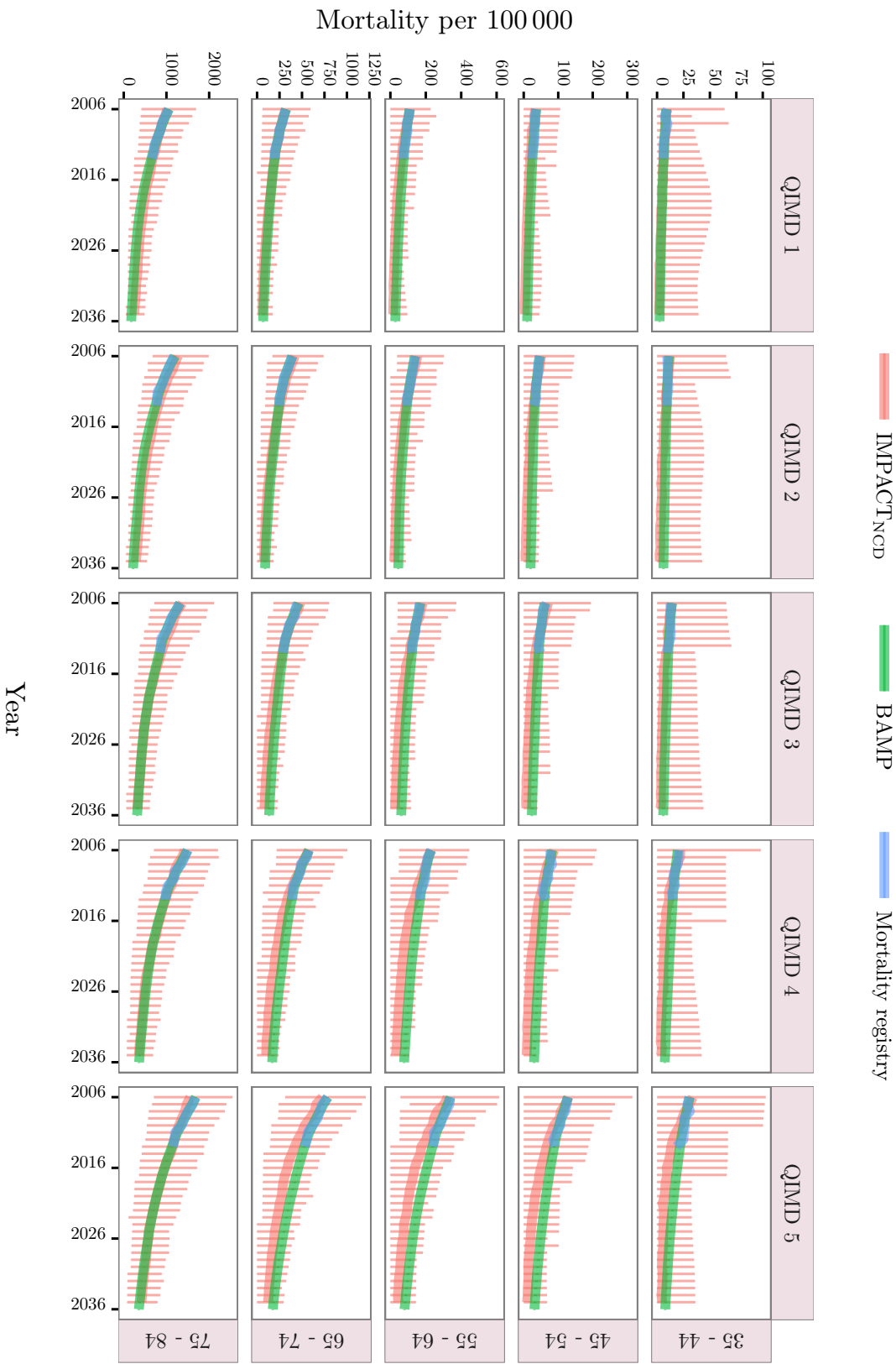
Figure 3.6: Men coronary heart disease mortality rate trends for ages 30 – 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.
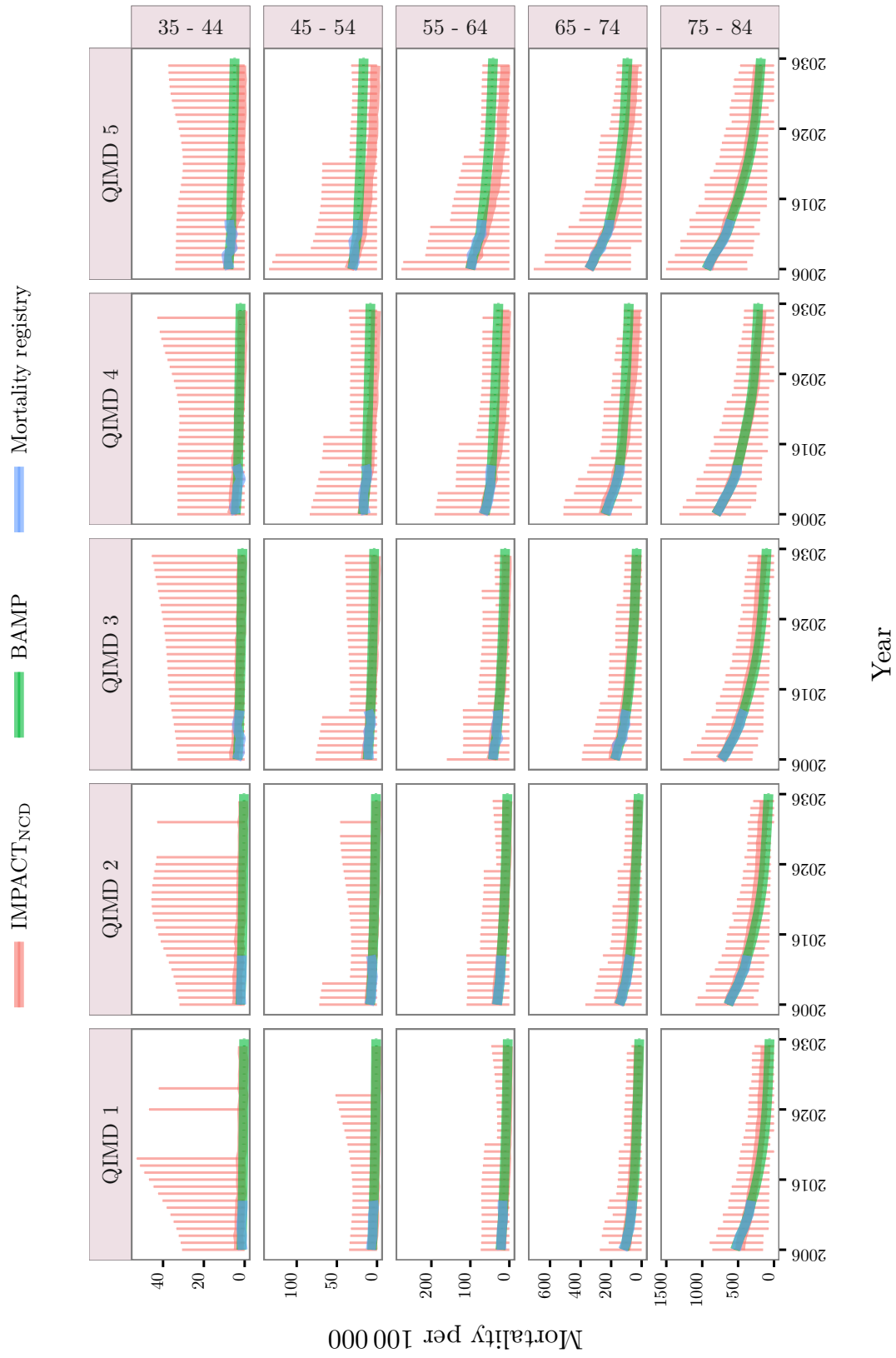
Figure 3.7: Women coronary heart disease mortality rate trends for ages 30 – 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.
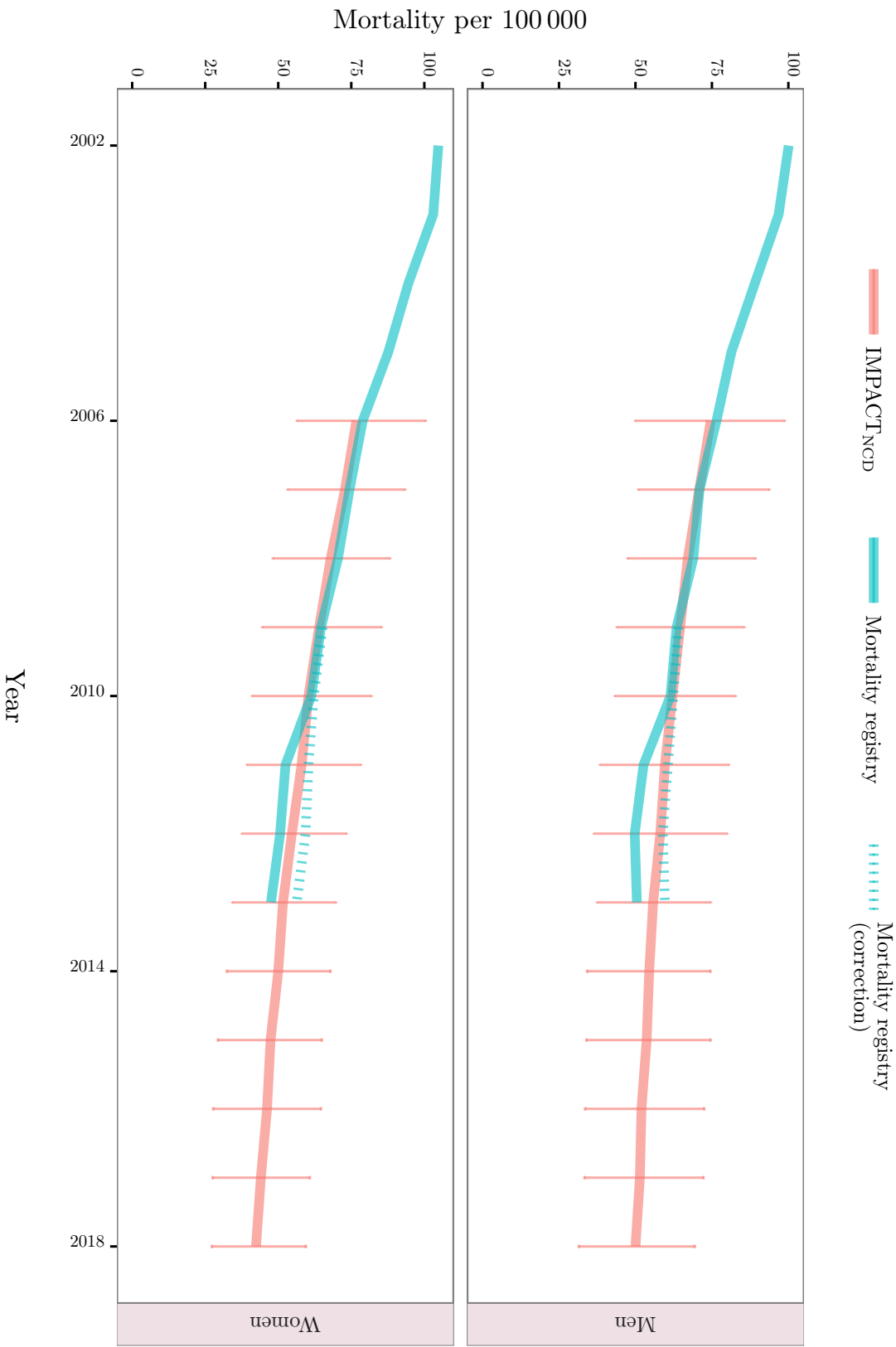
Figure 3.8: Stroke mortality rate trends for ages 30 – 84 between years 2002 and 2018. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. The corrected mortality (dotted line) adjusts to account for changes in the coding of vascular dementia since 2011.[297] Error bars depict 95 % uncertainty intervals.
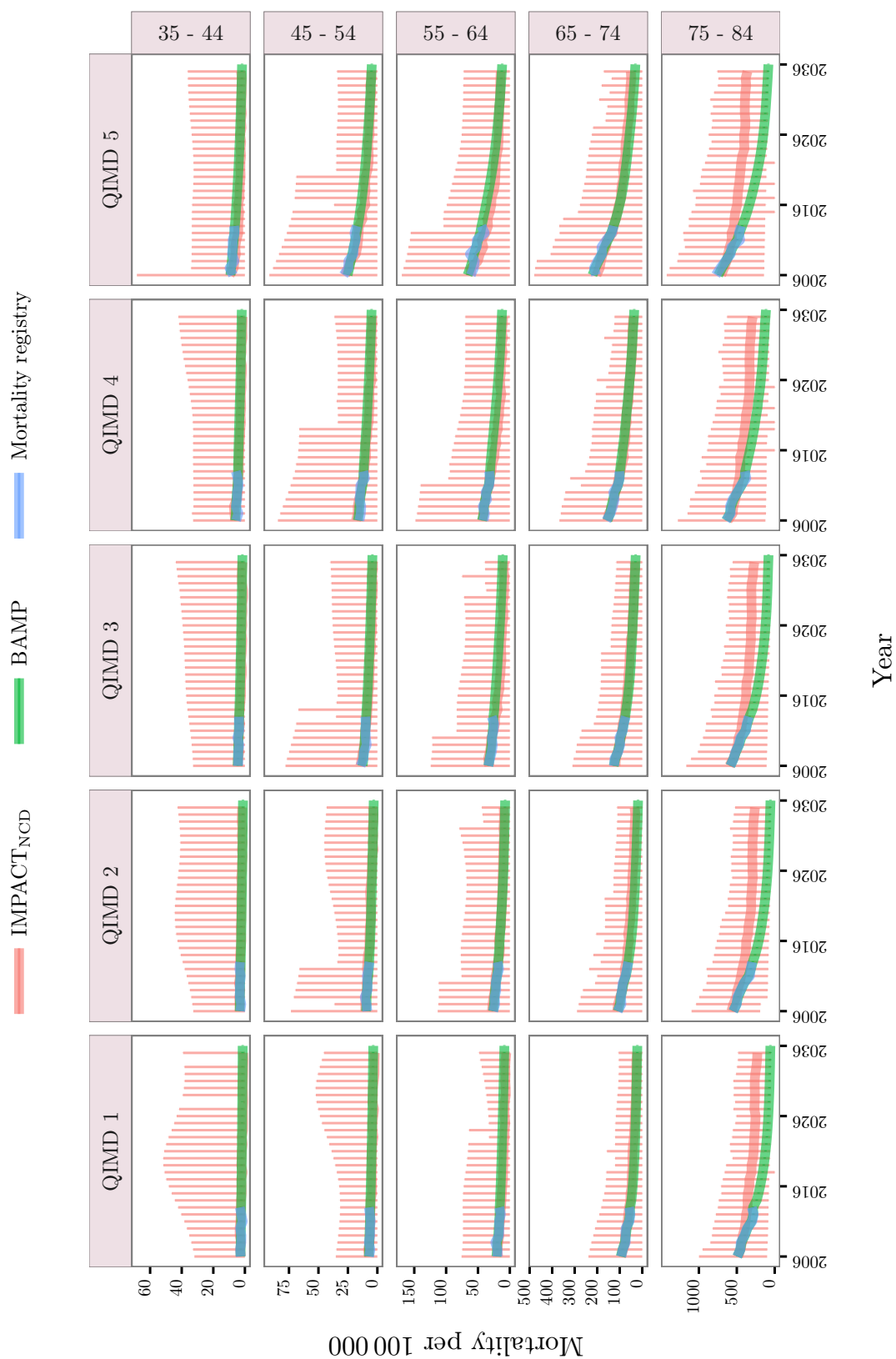
Figure 3.9: Men stroke mortality rate trends for ages 30 – 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.
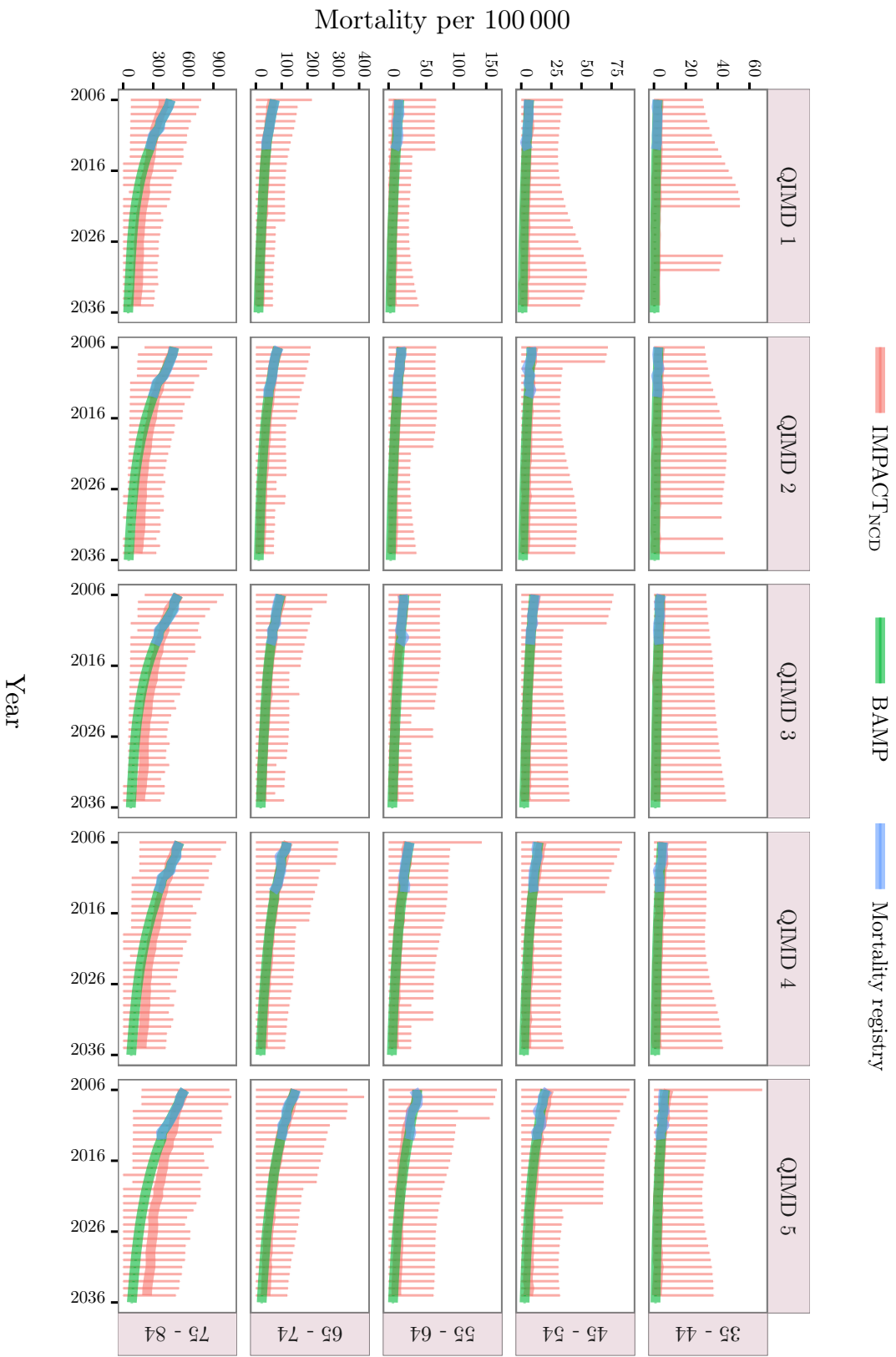
Figure 3.10: Women stroke mortality rate trends for ages 30 – 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.
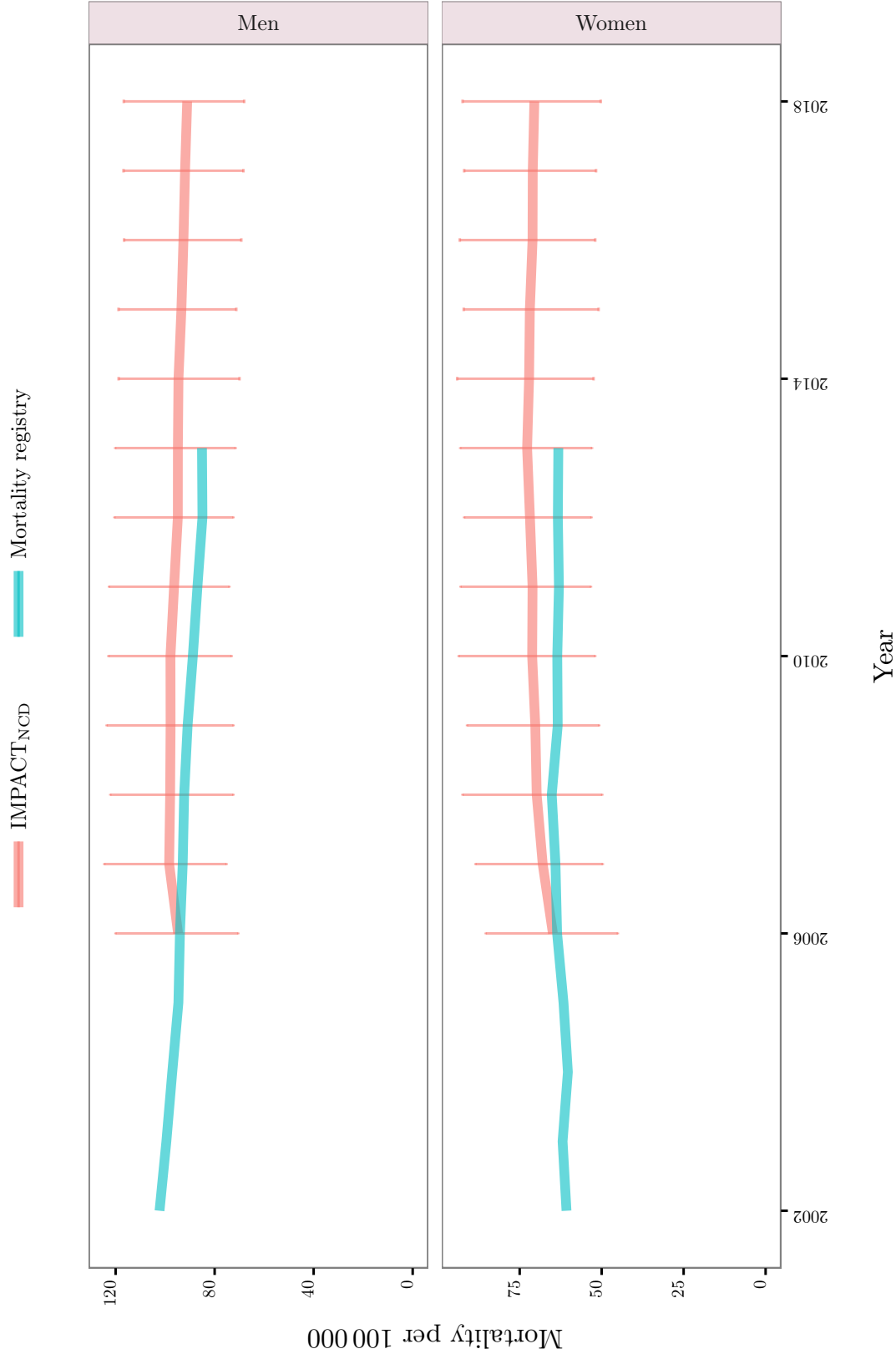
Figure 3.11: Lung cancer mortality rate trends for ages 30 − 84 between years 2002 and 2018. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % uncertainty intervals.
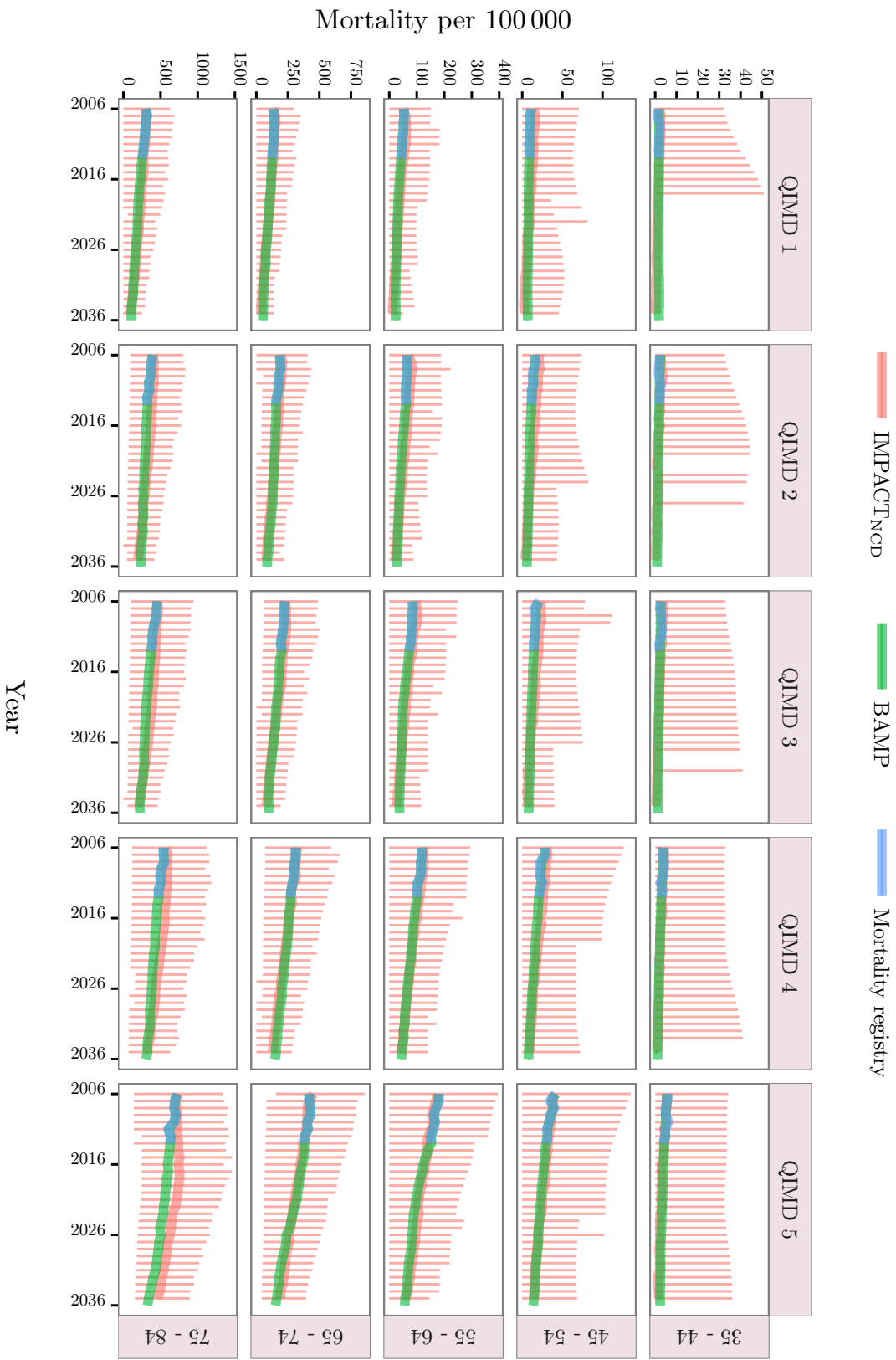
Figure 3.12: Men lung cancer mortality rate trends for ages 30 – 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.
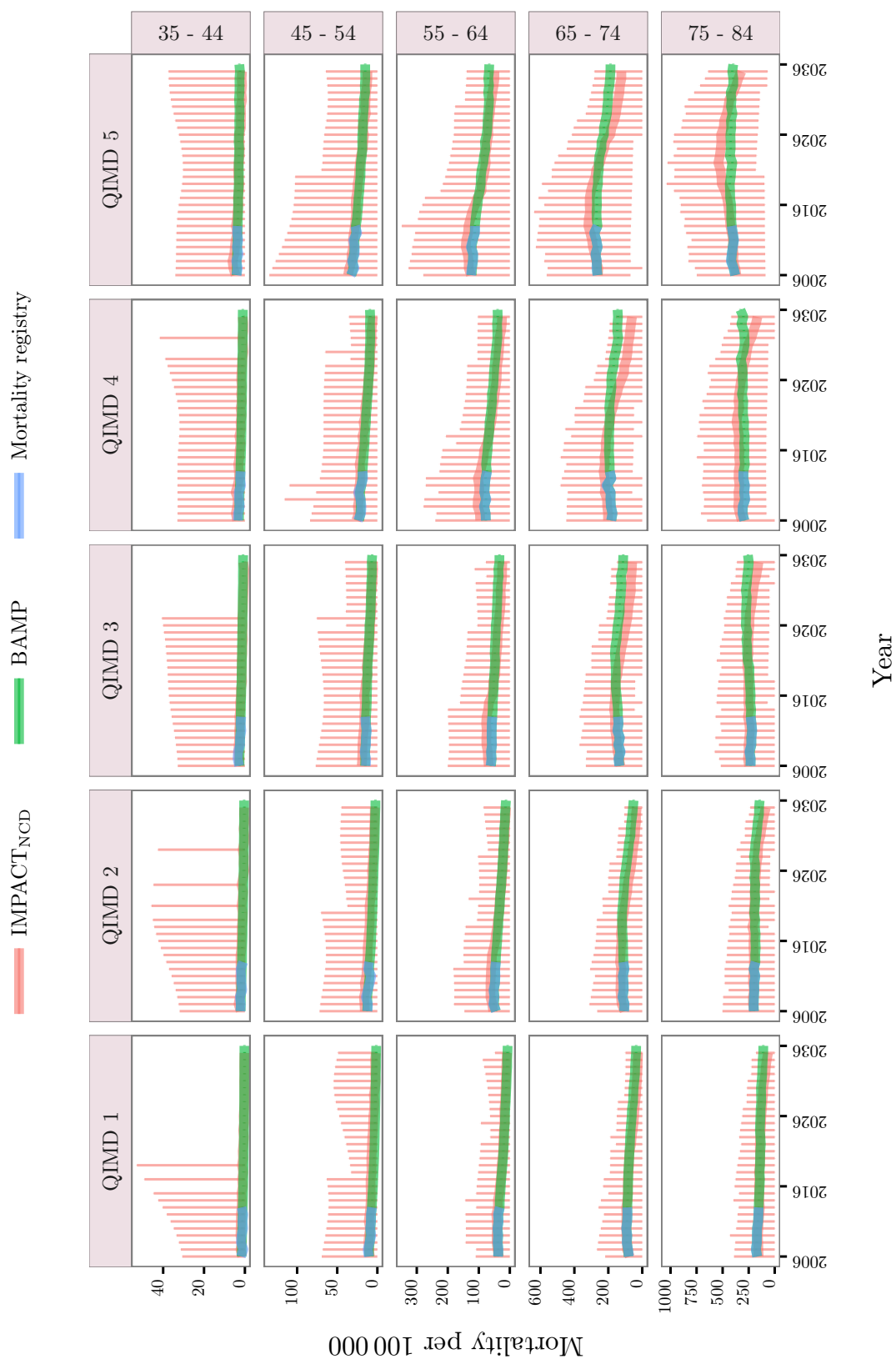
Figure 3.13: Women lung cancer mortality rate trends for ages 30 – 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.
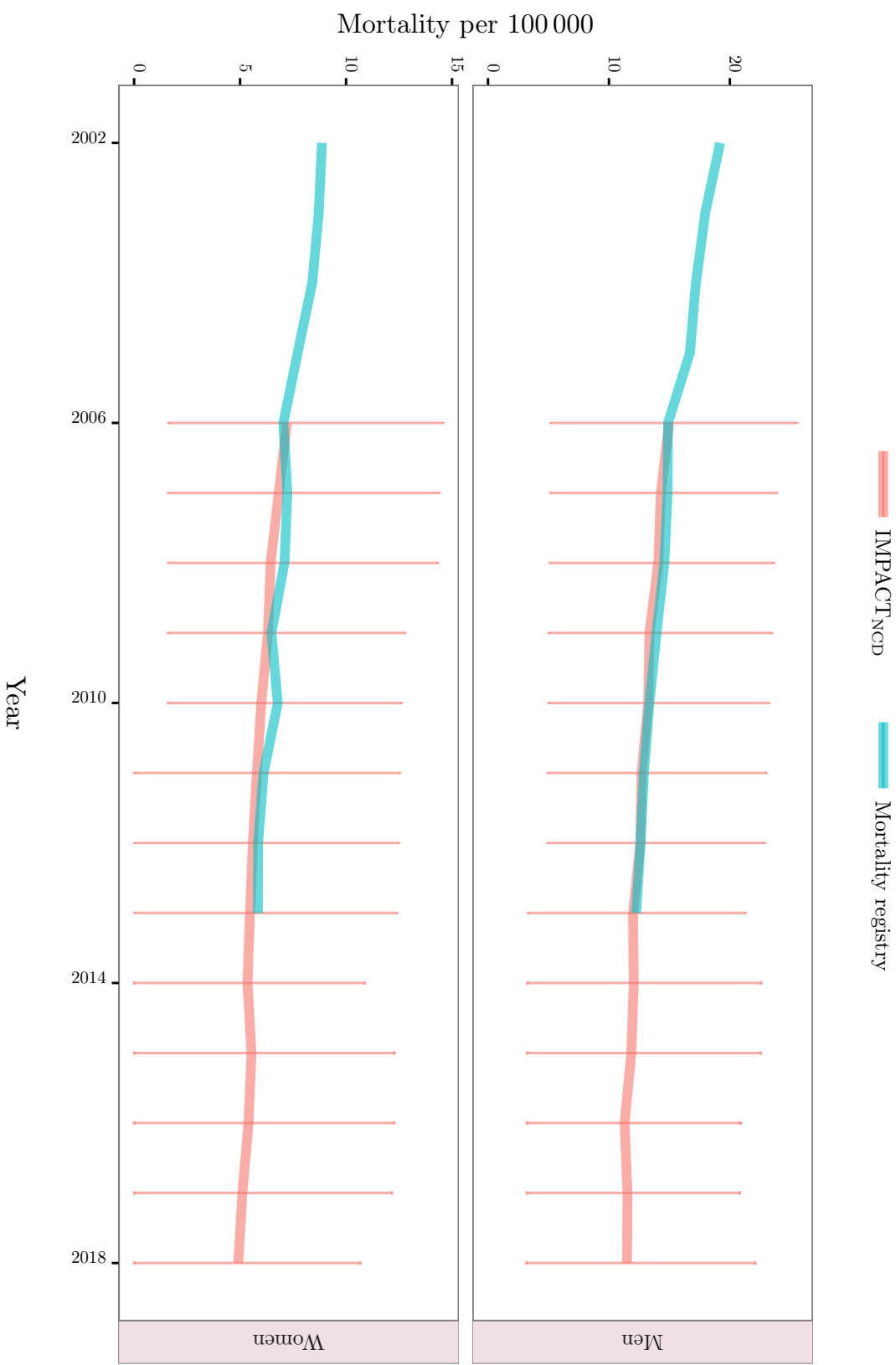
Figure 3.14: Gastric cancer mortality rate trends for ages 30 − 84 between years 2002 and 2018. Observed in the population through the mortality registry versus IMPACT$_{\text{NCD}}$ synthetic population estimates, stratified by sex. Error bars depict 95 % uncertainty intervals.
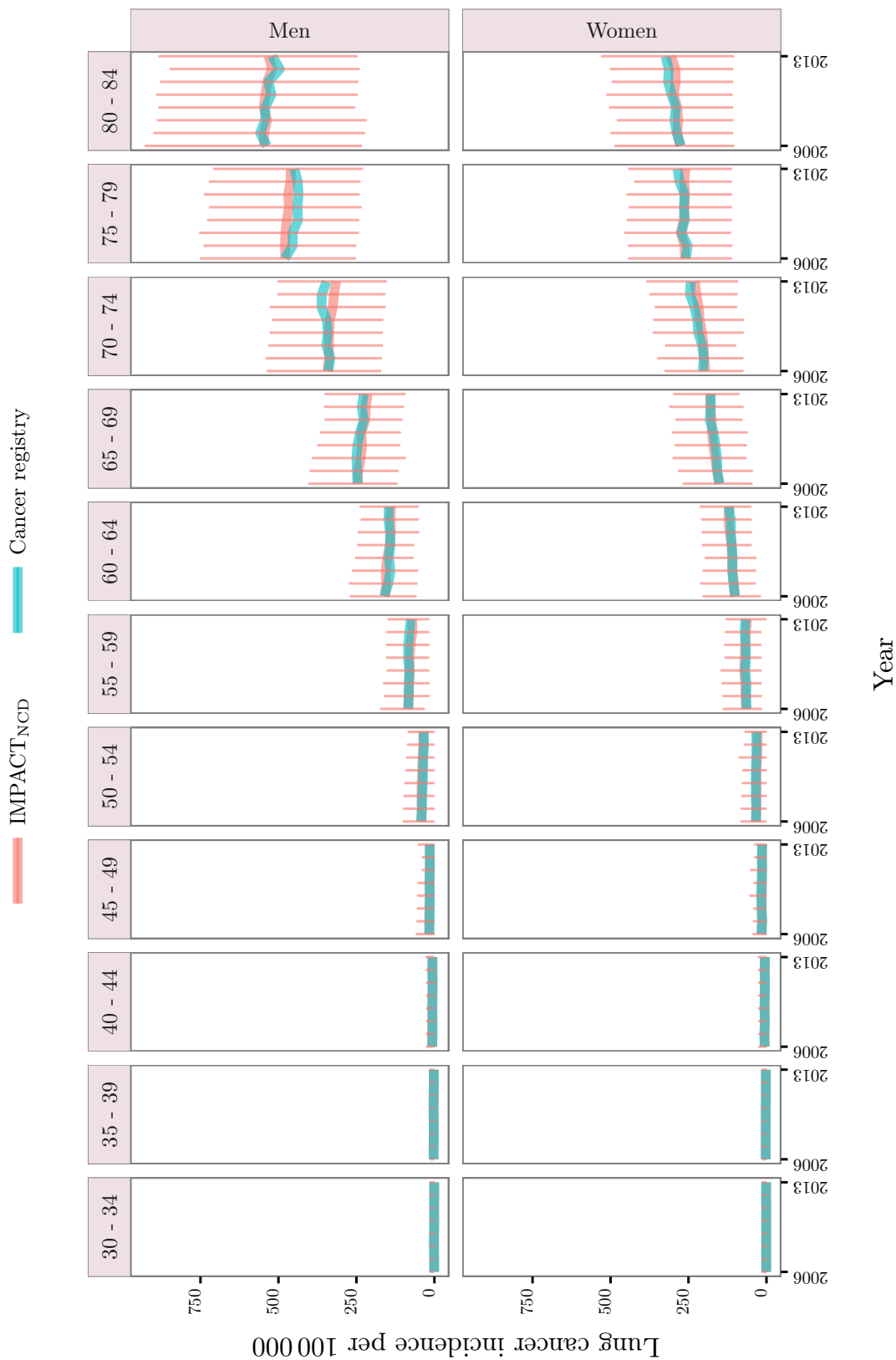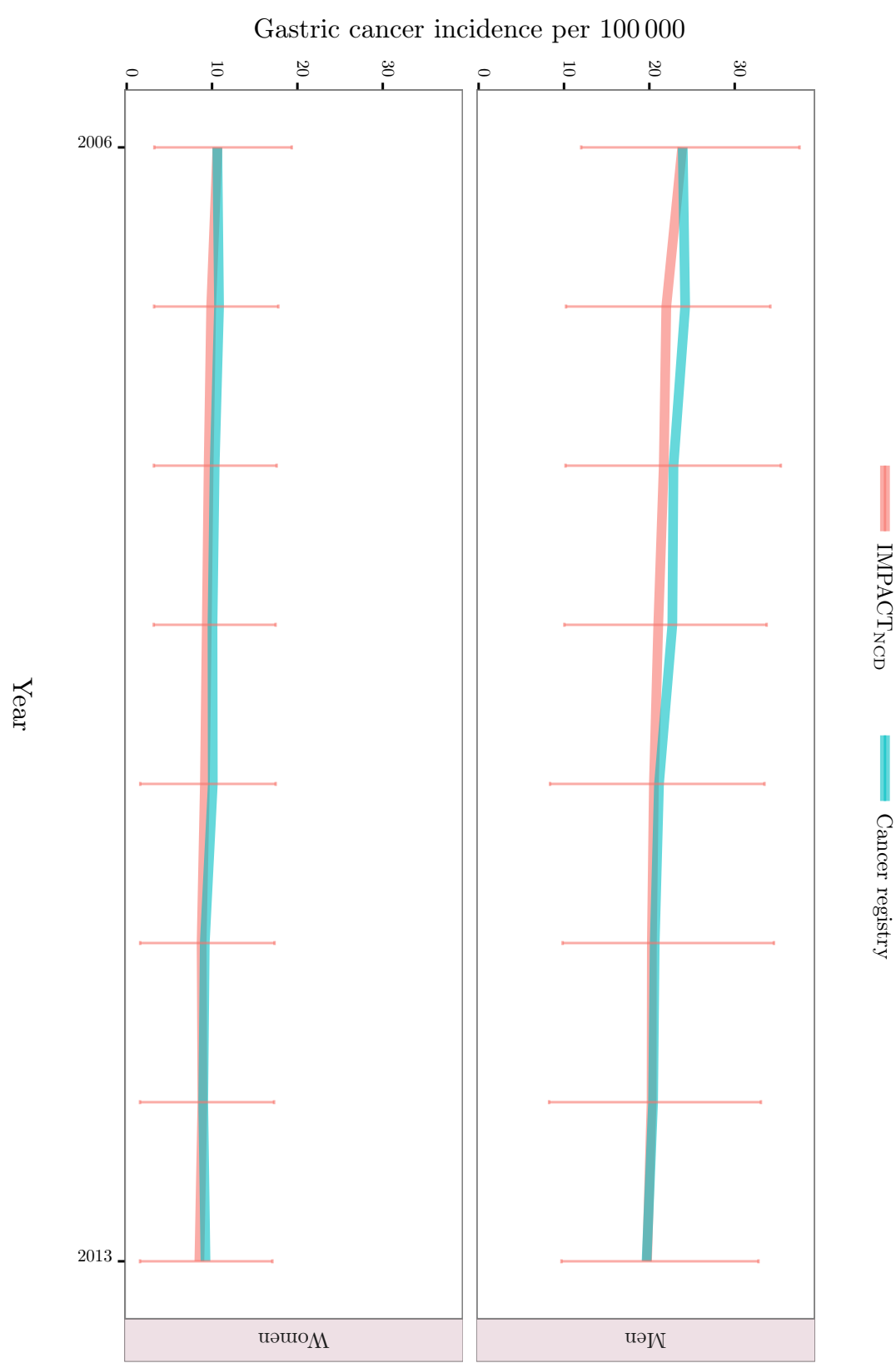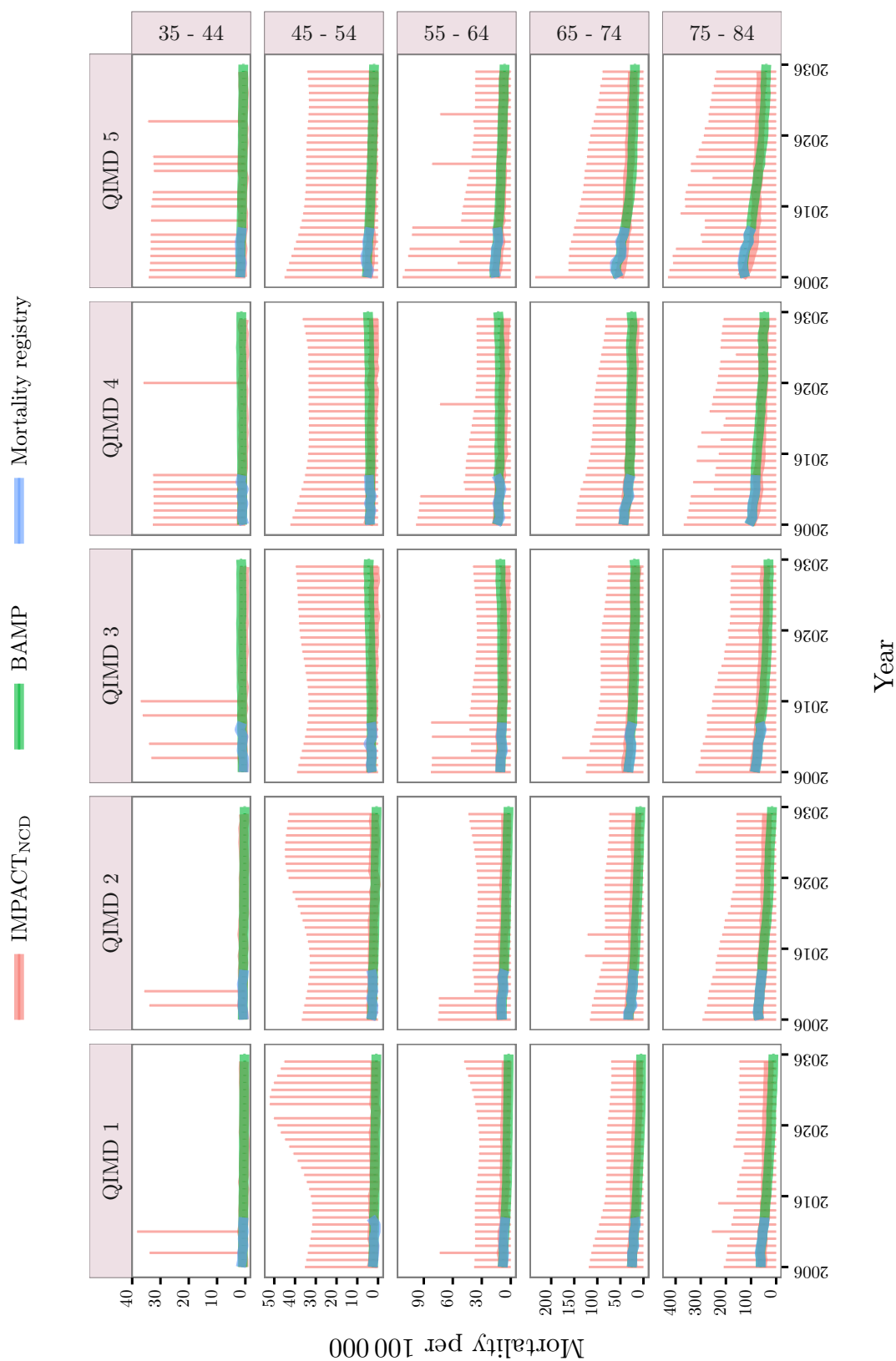
Figure 3.15: Men gastric cancer mortality rate trends for ages 30 − 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.
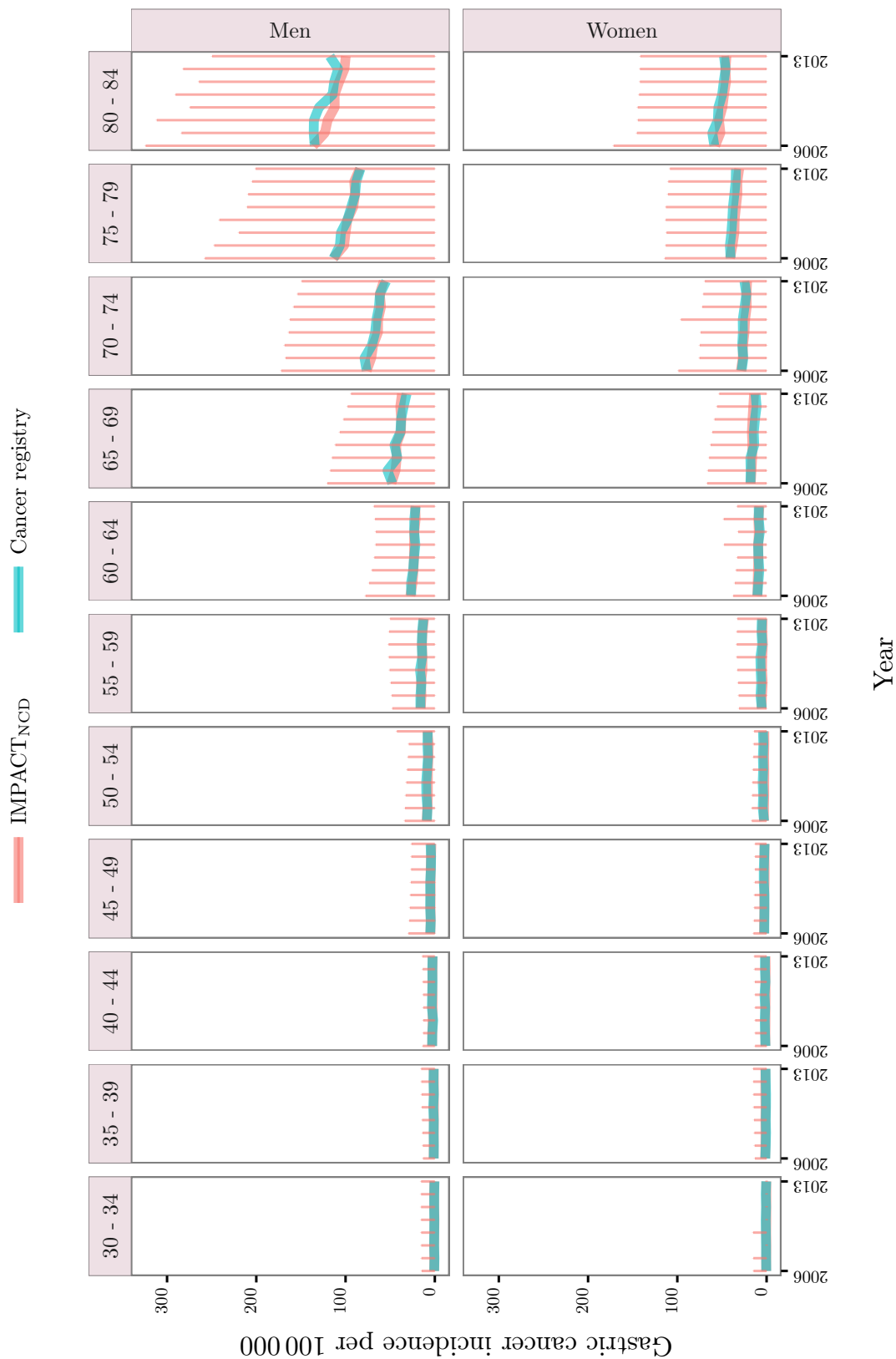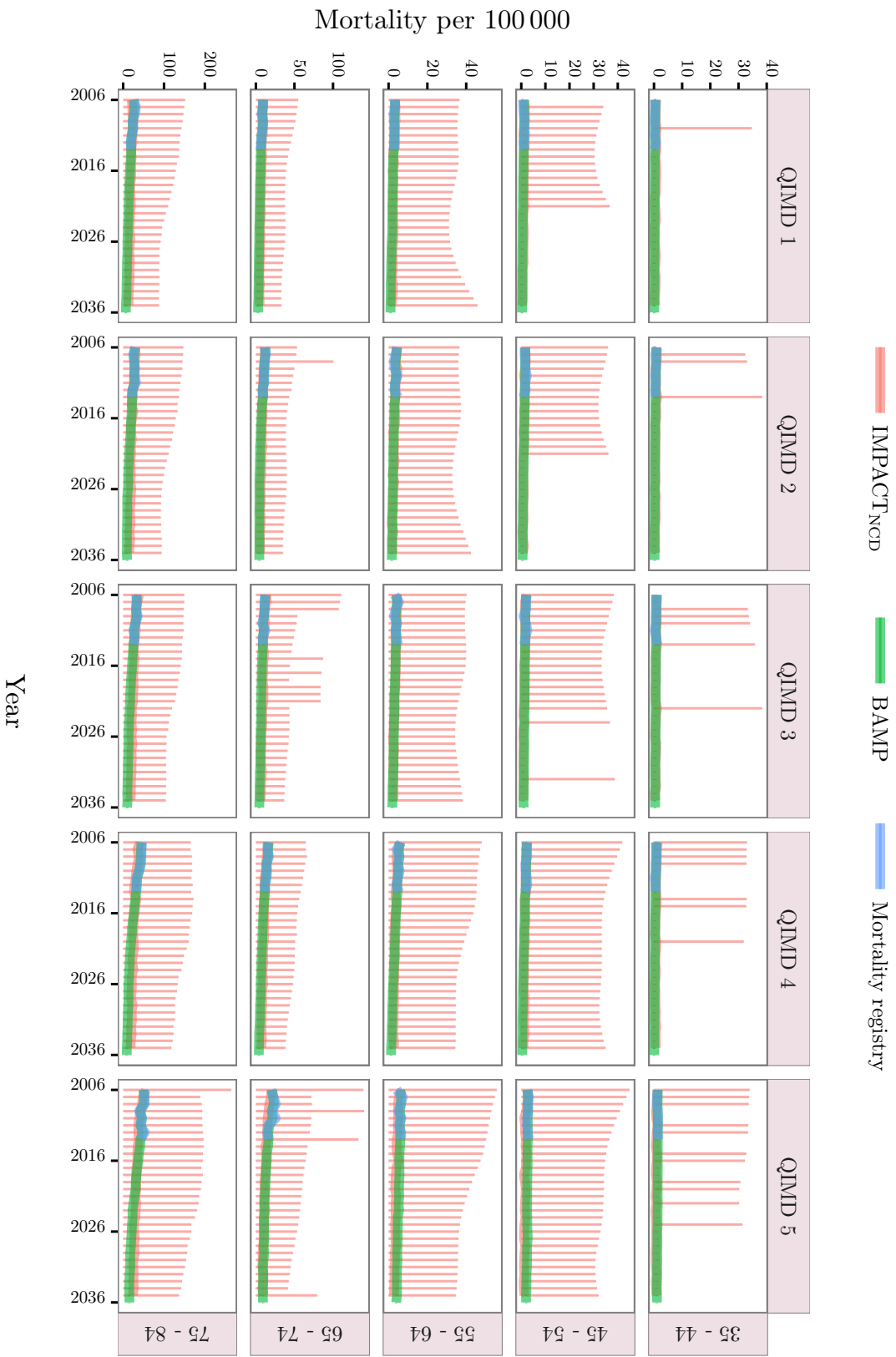
Figure 3.16: Women gastric cancer mortality rate trends for ages 30 – 84 between years 2006 and 2035. Observed in the population through the mortality registry versus IMPACT$_{NCD}$ synthetic population estimates versus the Bayesian age-period-cohort forecasts,[293] stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and age group. Error bars depict 95 % uncertainty intervals.

3.4 DISCUSSION

The previous graphs suggest that the synthetic world of IMPACT$_{NCD}$ replicates the real world reasonably well. They also present in a transparent and non-technical manner possible biases in the model that may be helpful when assessing model outputs. For example, the impact of smoke free places legislation led to a sharp decrease in the prevalence of environmental tobacco smoking between 2005 and 2008. I chose to model environmental tobacco smoking as a linear function of smoking prevalence for simplicity and computational efficiency. The impact of this assumption is explicitly depicted in the graph (figure B.22 on page 220), although the overall impact on disease incidence and mortality is rather small because environmental tobacco smoking has a small excess risk overall (table A.1 on page 187). The explicit nature of most modelling assumptions is an enormous advantage of modelling compared to all the traditional research methods.[156] In the following paragraphs I will discuss all the validation types for IMPACT$_{NCD}$.

3.4.1 *Face validity*

The development of IMPACT$_{NCD}$ was supervised by two NCD modelling experts, (Dr Martin O'Flaherty and Prof Simon Capewell), and a computational statistics and health informatics expert (Prof Iain Buchan). It was also scrutinised by co-authors and peer reviewers and published in a high impact medical journal with further manuscripts already submitted or pending submission.[298] In addition, the methods of the model have been presented and discussed with experts in Farr institute local meetings. The outputs of the model have been presented in several epidemiological conferences and local and national stakeholder meetings.[299–301] Finally, IMPACT$_{NCD}$ is scheduled to be used to model populations of countries other than England, as well as local populations within England in the near future.

3.4.2 *Internal validity*

The graphs in the results section (section 3.3 on page 79) are the obvious internal validation of IMPACT$_{NCD}$. The initial synthetic population is similar to the primer survey; the simulated life histories are replicating acceptably the recent observed population trends of risk factor exposures; and disease incidence and mortality for 2006 that were used as inputs are replicated in the outputs. However, the open-source paradigm of IMPACT$_{NCD}$ is a less obvious component of its internal validation. Apart from the IMPACT$_{NCD}$ source code itself that is open and can be reviewed and improved by the online community, all the R packages that IMPACT$_{NCD}$ uses are also open-source and are under scrutiny by experts from industry and academia. Moreover, R is an open-source programming language by statisticians and for statisticians.[209, 302] This ensures that all operations in R are naturally grounded in statistical methodology, unlike any other programming language.

3.4.3 *Cross validity*

I assessed the cross validity of IMPACT$_{NCD}$ by comparing its mortality rate estimates with a Bayesian age-period-cohort model.[293] The two model outputs are in agreement generally. Interestingly, during the validation process IMPACT$_{NCD}$ revealed that the sharp decrease of stroke deaths after 2010 was an artefact. On the contrary, the Bayesian age-period-cohort model was trapped in the data and projected the artefactual drop in the future. This vividly highlight the benefits of building models around core epidemiological principles. I further compare specific scenario outputs of IMPACT$_{NCD}$ with other model outputs in the following chapters, where I model specific policies.

3.4.4 *External and predictive validity*

IMPACT$_{NCD}$ incidence and mortality estimates for the modelled diseases have been externally validated against the ONS reported estimates. Given the aims and objectives of my thesis (section 1.7 on page 40) validation of the disease module was important and suggests that IMPACT$_{NCD}$ estimates are not far from reality. Furthermore, it shows that the model can successfully translate risk factor changes into incidence and then mortality changes. It is also able to replicate the socioeconomic gradients in risk factor exposures (differential exposure, figure 1.8 on page 27), the clustering of risk factor exposures in the most deprived groups (differential vulnerability, figure 1.8 on page 27), and the socioeconomic gradients in case fatalities (differential consequences, figure 1.8 on page 27). This is particularly important, because it essentially validates the underlying mechanism that is used in all modelled scenarios, later in the thesis. IMPACT$_{NCD}$ only exists for less than two years. Therefore, predictive validation would be impossible. It would be very interesting however, to compare the reported model outputs in this thesis with observed CVD and cancer incidence and mortality in the years to come.

3.5 CONCLUSIONS

Despite the numerous assumptions and limitations, this extensive validation exercise provides evidence that IMPACT$_{NCD}$ is 'valid enough' to support decision-making about primary prevention policies spanning populations and life courses. It is worth noting that all model inputs are informed by available data and epidemiological reasoning, and are not a product of an optimisation process or extensive calibration. Unfortunately, it is very rare for existing models to report an extensive validation process. In fact, the majority of published models report some fragments of validation, and many do not report any validation process. This may be one of the reasons modelling is sometimes treated with suspicion by public health policy makers and researchers. Yet, modelling provides the option for extensive validation and calibration of the modelling assumptions in comparison to traditional research methods that may use implicit, and therefore untested, assumptions.

This chapter concludes the first part of the thesis. In the next part I will use IMPACT$_{NCD}$ to model specific primary prevention policies and I will discuss the overarching themes that will emerge from these simulation exercises.

Part II

APPLICATIONS OF IMPACT$_{\text{NCD}}$

4

THE ROLE OF STATINS IN THE OBSERVED CHOLESTEROL DECLINE

4.1 INTRODUCTION

In chapter 1 I described that high serum total cholesterol is a risk factor for CVD and that the population exposure to this risk factor has been declining for some decades (section 1.3.8 on page 23). Consequently, serum total cholesterol remains one of the main targets for primary and secondary prevention of CVD and there is an ongoing discussion regarding what drives the observed decline. It is likely that initial decline was the result of dietary changes alone, in particular, the substitution of animal fats with vegetable oils in the English diet.[303] Nevertheless, over the last two decades the use of lipid lowering medication has increased dramatically in England. The raise can be attributed to a class of lipid lowering medicines known collectively as 'statins' that revolutionised the treatment options for hypercholesterolaemia.[304] Hence, more recent changes of mean cholesterol in the population are probably reflecting the interplay between improving diet and increasing statin use.

Despite a plethora of information on the effectiveness of statins at the individual level, especially for secondary prevention, their contribution to the total cholesterol fall in the wider population remains unclear. Farzadfar *et al.* and Cohen *et al.* suggest that statins are important in lowering population mean total cholesterol in high income countries including the US.[305, 306] However, it seems that this is neither completely true, nor universal because: 1. large falls in total cholesterol occurred before statins were widely used; [134, 303, 307, 308] and 2. the large recent total cholesterol falls observed in Iceland, Sweden, the Czech Republic, and Finland are principally attributed to improved diets.[309–312]

Unlike many other NCD risk factors, total cholesterol shows no socioeconomic gradient in young adults and an inverse socioeconomic gradient may exist for older adults, thus more affluent groups appear to have higher total cholesterol levels, especially since 1998.[83, 313] Given the known socioeconomic gradient in unhealthy diet in England,[88] it is not clear why no such gradient is observed in total cholesterol and what is the role of statins in this. Theoretically, statin prescription is a healthcare based intervention requiring individual action, which might potentially increase inequalities.[143, 314] Essentially, promoting statins for primary prevention of CVD is a typical high-risk, agentic intervention (please refer to section 1.5 on page 30 for an overview of policy typologies); yet, there is no obvious socioeconomic gradient in mean total cholesterol levels.

The debate about statins for primary prevention of CVD became heated a few years ago. In 2013, the American College of Cardiology and the American Heart Association updated their recommendations for the treatment of total cholesterol, substantially widening the criteria for statin prescription in otherwise healthy individuals.[315] A year later, NICE

made similar recommendations for England. Specifically, NICE now recommends statins prescription for healthy individuals that have more than a 10 % 10-year risk of developing CVD, down from 20 %.[316] This has almost doubled the number of eligible adults, from 7 to 12 million and has proved very controversial.[317, 318]

The primary objective of this chapter is to quantify the contribution of statins to the observed fall in population mean cholesterol levels in England over the past two decades. A secondary objective was to look for any differences in this contribution between socioeconomic groups.

4.2   METHODS

I used an early version of IMPACT$_{NCD}$ for this chapter. Therefore, many features of the fully developed model that have been described in the Methods (section 2.2 on page 43) were not yet mature enough at that time. Namely, the close to reality synthetic population and the dynamic traits of the synthetic individuals. Hence, in a static microsimulation framework (section 2.1.2 on page 43), I analysed changes in observed mean total cholesterol levels in England's adult population between 1991-92 (baseline) and 2011-12. I then compared the observed changes with a counterfactual 'no statins' scenario, where the impact of statins on population total cholesterol was estimated and removed. Any gap would then be attributed to all other possible drivers of population cholesterol levels, principally diet. I stratified my analysis by age group, sex, and where possible and relevant, by the 2010 version of QIMD.[319]

4.2.1   *Survey data*

In detail, I used individual level data from the HSE for the two respective periods.[204, 205, 320] For the 2011-12 period I aggregated the data of HSE2011 and HSE2012, while for 1991-92 this was independently performed by HSE analysts. These cross-sectional surveys provide a representative sample of the community dwelling population in England for the respective years (please refer to section 2.3.1 on page 46 for a more detailed description). The data files contained anonymised information for all the participants. I excluded participants younger than 18 years old. For HSE2011-12 both the weighting and the sampling design were considered in the estimation of all the point estimates and their standard errors. In particular, the weighting adjusts both for selection and non-response bias. The sample for HSE1991-92 was unweighted, therefore, only the sampling design was taken into account.

4.2.2   *Socioeconomic stratification*

Unfortunately, there were no common socioeconomic indicators between the two samples; QIMD was therefore used for the 2011-12 sample and social class based on occupation (I – V) was used for the 1991-92 sample. The HSE team provided the QIMD of each participant

for HSE2011-12 based on their postcode of residence. I opted to use the QIMD instead of other available socioeconomic classification systems mainly for two reasons. First, the QIMD was the only socioeconomic indicator that had no missing cases in my data; and second, for my results to be comparable with other studies that used QIMD. The HSE1991-92 social class classification was based on the 1990 version of the Standard Occupational Classification and the self-reported occupation of the participants.[94] Social class was provided as a variable in the data, by the HSE team. I aggregated full time students, armed forces personnel, those who never worked, and those whose occupation was not fully described in one category ('Other'). In my analysis, I avoided any direct comparisons between the two socioeconomic classification systems.

4.2.3 *Total cholesterol measurement*

In 2011-12 a smaller sample within the total HSE sample was eligible and consented to provide non-fasting blood samples for the measurement of total cholesterol in serum. For HSE1991-92, participants aged 18 and over were asked to provide a blood sample for the same purpose. Since April 2010 the equipment that was used for the measurement of total cholesterol for HSE was replaced. The effect of this change was that measured concentrations of total cholesterol from this date onwards were on average 0.1 mmol/l higher. I adjusted for this difference in my analysis by subtracting 0.1 mmol/l from the respective total cholesterol measurements.

4.2.4 *Estimating statin utilisation*

In England, individuals may have access to statins using two available routes. Statins can either be prescribed to them by a doctor (or an authorised non-medical prescriber), or they can be bought over the counter from a pharmacy with or without prior expert advice. HSE assessed both routes. In 2011-12, during a nurse interview, the participants were asked to report the medication that had been prescribed to them by a doctor or by a non-medical prescriber. Specifically for statins, they were also asked whether these were bought over the counter. Finally, those that had been prescribed a statin or bought it over the counter were asked if they had used it during the past seven days. I only considered the participants that answered positively in the last question as statin users. For HSE1991-92 the participants were asked similar questions during the nurse interview. However, statins were included in the wider category of lipid lowering medication and not recorded separately as in 2011-12. In any case, statin prescription was not prevalent before 2000 and it was primarily limited for secondary prevention.[308, 321] Since the uptake of the lipid lowering medication category as a whole was very low, I assumed that statins had a negligible effect on total cholesterol at population level; thus, I ignored it completely. For further justification of this assumption please refer to appendix C.1 on page 245.

4.2.5   *Statistical analysis*

The analysis was performed in R statistical software and the R package 'survey'.[209, 230] An approximate 95 % confidence interval (CI) for proportions (e. g. statin uptake) was calculated from the survey data using the incomplete beta function method, with an effective sample size based on the estimated variance of the proportion.[322] Missing cases were excluded from the analysis.

To test the statistical significance of socioeconomic trends in total cholesterol, against the null hypothesis of 'no trend', I fitted a generalised linear model, with inverse probability weighting and design based standard error (SE). Specifically, I treated total cholesterol measurements as the dependent variable and the QIMD (or social class) as the independent one. I considered QIMD and social class as numeric variables for this (e. g. QIMD 1 through 5 represented the 5 quintiles and social class 1 through 7 represented the social classes I, II, IIIN, IIIM, IV, V and 'Other', respectively). Therefore, the β coefficient (slope) of the QIMD (or social class) and its SE was a measure of the socioeconomic gradient. When β was not statistically significant I assumed no socioeconomic gradient. When β was statistically significant, its sign revealed the direction of the gradient (e. g. a negative sign means that mean total cholesterol is lower among the more deprived groups) and its absolute value measured the magnitude of the gradient. A similar approach was followed to explore socioeconomic trends in statin utilisation. Since this time the dependent variable was a binary one, I used a binomial model.

4.2.5.1   *Estimating the effect of statins*

The average effect of each specific statin and strength on an individual's total cholesterol is known from the literature.[323–326] However, the exact type of statin and strength had not been recorded for the participants in HSE2011-12. To overcome this limitation I used the exact amount of statins (by proprietary name and strength) that were prescribed and dispensed in England for 2011 and 2012,[327, 328] to estimate a weighted mean of the proportional decrease of total cholesterol attributable to statins overall using the equation:

$$E_w = \frac{\sum_i \sum_j (M_{ij} E_{ij})}{\sum_i \sum_j (M_{ij})}, \tag{4.1}$$

Where

E_w the proportional decrease in mean total cholesterol attributable to statins, among statin users

E_{ij} the proportional decrease in mean total cholesterol attributable to a specific statin i of a specific strength j (e. g. Simvastatin 20 mg)

M_{ij} the number of units of a specific statin i and strength j that have been prescribed and dispensed[37]

---

37 A unit was considered equivalent of a tablet, or 5 ml for liquid forms.

For the estimation of E_{ij} data from four meta-analyses were used as follows: I obtained the mean and SE (calculated directly from the 95 % CI assuming approximate normality) of the proportional reduction in serum low density lipoprotein (LDL) from the meta-analysis of Law *et al.* The proportional reduction was derived from the absolute reduction, standardised to usual serum LDL of 4.8 mmol/l before treatment and it was independent of the pre-treatment LDL.[329] This allowed me to use a weighted mean approach on proportions.

I then converted the LDL reduction into total cholesterol reduction using data from other studies,[324–326] assuming an approximately linear relation between total cholesterol and LDL reduction. For strengths not included in the above meta-analyses (e. g. Atorvastatin 30 mg), I used a log-linear regression model to estimate their effect, based on the effect of known strengths. I weighted the model against the inverse variance of the cholesterol reduction. I considered the effectiveness of solid and liquid forms being approximately equal. Similarly, the effectiveness of the combined forms of simvastatin with ezetimibe was considered equal to the effectiveness of same strength simvastatin (table C.1 on page 245). The SE of E_w was estimated using Cochran's definition for the SE of the weighted mean.[330, 331]

For the no statins scenario, I calculated the predicted total cholesterol for each statin user, with the statin effect removed using the equation:

$$TC_{pred} = \frac{TC_{obs}}{1 - E_w},$$ (4.2)

Where

TC_{pred} the predicted total cholesterol of the statin user with the statin effect removed

TC_{obs} the observed total cholesterol of the statin user

E_w the proportional decrease in mean total cholesterol attributable to statins, derived from equation (4.1) on page 106.

I used Monte Carlo simulation to propagate the uncertainty from the sampling distribution of E_w. For each statin user I drew 1000 values from a normal distribution with mean E_w and standard deviation (SD) as per the estimated SE (described above). I then averaged over the TC_{pred} predictions and considered this mean value as the predicted total cholesterol of each statin user, with the statin effect removed.

4.2.5.2 *Quantifying the contribution of statins on the population's mean total cholesterol reduction*

To quantify and compare the contribution of statins against the contribution of all other total cholesterol lowering interventions in the population, I first plotted the mean total cholesterol for 1991-92, 2011-12 and the no statins scenario by age for each sex. I considered the area enclosed by the respective curves for 1991-92 and 2011-12 as representing the full observed cholesterol reduction (area A). Therefore, the area enclosed by the 2011-12 and the 'no statin' scenario represents the reduction of cholesterol attributable to statins (area B). Thus, the fraction (area B) / (area A) expresses the contribution of statins to the

observed decline of mean total cholesterol. For the estimation of areas A and B, I used natural spline interpolation as implemented in the R package 'MESS'.[332]

To estimate the UI around the estimated contribution of statins, I modified the previous method to allow for a Monte Carlo simulation approach. Specifically, for each age in the population, I drew 10 000 values from the conditional sampling distribution, which I approximated by a normal distribution with age-specific estimate mean and SE. These were then averaged across the age range to yield a point estimate, and 2.5 % and 97.5 % percentiles were used to define the 95 % UI. Due to the small representation of ages above 89 in the sample, I aggregated participants older than 89 years with those aged 89.

Finally, I repeated the analysis separately for each QIMD under the assumption that total cholesterol had no socioeconomic gradient in 1991-92. I further limited the analysis in participants younger than 76 years because of the small number of older participants in the sample, when stratified by QIMD. To test the statistical significance of any observed socioeconomic trend I used the two tailed Cochran-Armitage trend test.[333]

4.2.6   *Sensitivity analysis*

For the estimation of E_w several assumptions were involved that do not necessarily reflect on its estimated SE. I repeated my analysis after I multiplied the standard error of E_w by a factor of 10 in order to test the robustness of my results with a higher than measured uncertainty scenario.

4.3   RESULTS

The baseline characteristics of the 1991-92 and 2011-12 samples are summarised in table 4.1 on the facing page, while mean total cholesterol values by age group and sex are presented in table 4.2 on page 110 (1991-92) and table 4.3 on page 111 (2011-12). Overall, the prevalence of statin use in England, including over the counter statin users was about 13.2 % (95 % CI: 12.5 % to 14.0 %) in 2011-12. Another about 0.8 % (95 % CI: 0.6 % to 1.0 %) of the population were prescribed or bought over the counter statins; however, they did not use them for at least a week before the nurse interview.

For 1991-92, statin use was not specifically recorded in the survey; however, the prevalence of all lipid lowering medications, including statins, was approximately 0.5 % (95 % CI: 0.3 % to 1.0 %). Table 4.4 on page 112 summarises the prevalence of statin use in England for 2011-12 by age group, sex, and QIMD. There was a statistically significant socioeconomic gradient in ages over 35 years for both sexes, in which the use of statins increased with deprivation.

In 2011-12, some 13.1 % (95 % CI: 12.4 % to 14.0 %) of study population used statins prescribed to them (not including over the counter users), over the seven days before the survey interview. I estimated the expected number of units (e. g. tablets or 5 ml doses of liquid statins) that were consumed in England for the same period, assuming that they stayed on statins for the whole year and that the institutionalised population shares the same consumption attitudes, to be approximately 4 billion. This showed reassuringly close agree-

Table 4.1: Samples baseline characteristics. Values are numbers (percentages).

| | Participants interviewed by a nurse | | | | Participants with a valid total cholesterol result | | | |
| | 1991-92 (n = 7043) | | 2011-12 (n = 10 965) | | 1991-92 (n = 4995) | | 2011-12 (n = 7772) | |
	Men	Women	Men	Women	Men	Women	Men	Women
Age (years)								
18 to 34	999 (14.2)	1165 (16.5)	877 (8.0)	1350 (12.3)	733 (14.7)	730 (14.6)	604 (7.8)	797 (10.3)
35 to 54	1148 (16.3)	1240 (17.6)	1632 (14.9)	2194 (20.0)	886 (17.7)	921 (18.4)	1216 (15.6)	1633 (21.0)
55+	1101 (15.6)	1390 (19.7)	2254 (19.7)	2658 (24.2)	806 (16.1)	919 (18.4)	1611 (20.7)	1911 (24.6)
Quintile groups of Index of Multiple Deprivation								
1 (least deprived)	–	–	1058 (9.6)	1389 (12.7)	–	–	785 (10.1)	995 (12.8)
2	–	–	1057 (9.6)	1364 (12.4)	–	–	791 (10.2)	997 (12.8)
3	–	–	1017 (9.3)	1278 (11.7)	–	–	732 (9.4)	892 (11.5)
4	–	–	865 (7.9)	1133 (10.3)	–	–	606 (7.8)	781 (10.0)
5 (most deprived)	–	–	766 (7.0)	1038 (9.5)	–	–	517 (6.7)	676 (8.7)
Social class								
I Professional	235 (3.3)	53 (0.8)	–	–	174 (3.5)	41 (0.8)	–	–
II Managerial technical	908 (12.9)	856 (12.2)	–	–	688 (13.8)	610 (12.2)	–	–
IIIN Skilled non-manual	320 (4.5)	1304 (18.5)	–	–	238 (4.8)	909 (18.2)	–	–
IIIM Skilled manual	1085 (15.4)	388 (5.5)	–	–	816 (16.3)	251 (5.0)	–	–
IV Semi skilled manual	460 (6.5)	693 (9.8)	–	–	343 (6.9)	464 (9.3)	–	–
V Unskilled manual	157 (2.2)	363 (5.2)	–	–	112 (2.2)	225 (4.5)	–	–
Other	83 (1.2)	138 (2.0)	–	–	54 (1.1)	70 (1.4)	–	–

The difference between the number of participants that had a nurse interview and those who had a valid total cholesterol result indicates the missing cases.

Table 4.2: Observed mean total cholesterol (mmol/l) overall, and by age group, sex and social class in England, 1991-92. Socioeconomic trends are also presented.

Social class	18 to 34 (years)		35 to 54		55+		Overall
	Men	Women	Men	Women	Men	Women	
I Professional	5.5 (5.2 to 5.8)	5.1 (4.7 to 5.5)	6.0 (5.7 to 6.2)	5.6 (5.3 to 6.0)	6.0 (5.7 to 6.3)	6.6 (6.1 to 7.1)	5.6 (5.5 to 5.8)
II Managerial technical	5.3 (5.1 to 5.4)	5.1 (4.9 to 5.2)	6.0 (5.9 to 6.1)	5.6 (5.5 to 5.7)	6.2 (6.1 to 6.4)	6.8 (6.6 to 7.0)	5.7 (5.6 to 5.8)
IIIN Skilled non-manual	5.2 (5.1 to 5.4)	5.0 (4.9 to 5.1)	6.2 (5.9 to 6.4)	5.7 (5.6 to 5.8)	6.1 (5.8 to 6.4)	6.8 (6.7 to 6.9)	5.6 (5.5 to 5.8)
IIIM Skilled manual	5.2 (5.0 to 5.3)	5.1 (4.8 to 5.3)	5.9 (5.8 to 6.1)	6.0 (5.7 to 6.2)	6.1 (6.0 to 6.2)	6.8 (6.6 to 7.1)	5.7 (5.6 to 5.8)
IV Semi skilled manual	5.2 (5.0 to 5.4)	5.1 (5.0 to 5.3)	5.9 (5.7 to 6.1)	5.7 (5.5 to 5.9)	6.0 (5.8 to 6.2)	7.0 (6.8 to 7.1)	5.7 (5.6 to 5.9)
V Unskilled manual	5.3 (4.8 to 5.7)	5.2 (4.8 to 5.5)	6.1 (5.7 to 6.5)	6.00 (5.8 to 6.2)	6.0 (5.6 to 6.5)	7.0 (6.5 to 7.4)	6.0 (5.8 to 6.2)
Other	4.7 (4.4 to 5.0)	5.1 (4.8 to 5.5)	5.8 (5.0 to 6.6)	5.0 (4.5 to 5.6)	6.4 (5.6 to 7.1)	6.7 (6.1 to 7.3)	5.3 (5.1 to 5.5)
All	5.2 (5.1 to 5.3)	5.1 (5.0 to 5.1)	6.0 (5.9 to 6.1)	5.7 (5.6 to 5.8)	6.1 (6.0 to 6.2)	6.8 (6.8 to 6.9)	5.7 (5.6 to 5.8)
Slope of the trend	-0.07 (-0.13 to -0.01)	0.02 (-0.02 to 0.07)	-0.02 (-0.07 to 0.03)	0.05 (0.00 to 0.10)	-0.04 (-0.10 to 0.02)	0.04 (-0.04 to 0.11)	0.00 (-0.02 to 0.02)*
p for trend	0.01	0.27	0.47	0.03	0.19	0.32	0.96*

* Adjusted for age and sex.
Brackets contain 95% confidence interval.

Table 4.3: Observed mean total cholesterol (mmol/l) overall, and by age group, sex and quintile groups of Index of Multiple Deprivation (QIMD) in England, 2011-12. Socioeconomic trends are also presented.

QIMD	18 to 34 (years)		35 to 54		55+		
	Men	Women	Men	Women	Men	Women	Overall
1 (least deprived)	4.8 (4.6 to 5.0)	4.8 (4.6-4.9)	5.5 (5.4 to 5.6)	5.2 (5.1 to 5.4)	5.1 (5.0 to 5.2)	5.8 (5.7 to 5.9)	5.2 (5.1 to 5.3)
2	4.7 (4.6 to 4.9)	4.5 (4.3 to 4.6)	5.5 (5.3 to 5.6)	5.2 (5.1 to 5.3)	5.1 (5.0 to 5.2)	5.7 (5.6 to 5.8)	5.1 (5.0 to 5.2)
3	4.6 (4.4 to 4.9)	4.7 (4.5 to 4.9)	5.6 (5.5 to 5.8)	5.3 (5.2 to 5.4)	5.1 (4.9 to 5.2)	5.7 (5.5 to 5.8)	5.1 (5.0 to 5.2)
4	4.8 (4.7 to 5.0)	4.6 (4.5 to 4.8)	5.5 (5.3 to 5.6)	5.4 (5.2 to 5.5)	5.0 (4.8 to 5.1)	5.6 (5.4 to 5.7)	5.1 (4.9 to 5.2)
5 (most deprived)	4.8 (4.6 to 5.0)	4.6 (4.4 to 4.7)	5.4 (5.2 to 5.6)	5.3 (5.2 to 5.5)	4.7 (4.6 to 4.9)	5.3 (5.2 to 5.5)	4.9 (4.8 to 5.1)
All	4.8 (4.7 to 4.8)	4.6 (4.6 to 4.7)	5.5 (5.4 to 5.6)	5.3 (5.2 to 5.3)	5.0 (5.0 to 5.1)	5.6 (5.6 to 5.70)	
Slope of the trend	0.02 (-0.05 to 0.08)	-0.01 (-0.06 to 0.04)	-0.03 (-0.07 to 0.02)	0.03 (-0.01 to 0.07)	-0.08 (-0.12 to -0.03)	-0.10 (-0.14 to -0.05)	-0.03 (-0.05 to -0.01)*
p for trend	0.60	0.67	0.26	0.16	<0.001	<0.001	0.002*

* Adjusted for age and sex.
Brackets contain 95% confidence interval.

Table 4.4: Prevalence of statin use in England 2011-12 by age, sex and quintile groups of Index of Multiple Deprivation (QIMD), 2011-12. Socioeconomic trends are also presented.

QIMD	18 to 34 (years)		35 to 54		55+		Overall
	Men	Women	Men	Women	Men	Women	
1 (most affluent)	0% (0 to 2%)	–	5% (3 to 8%)	2% (1 to 3%)	36% (32 to 41%)	20% (16 to 23%)	19% (16 to 23%)
2	–	–	7% (4 to 11%)	3% (2 to 5%)	38% (34 to 43%)	24% (20 to 27%)	22% (18 to 26%)
3	0% (0 to 2%)	0% (0 to 2%)	7% (5 to 11%)	2% (1 to 4%)	32% (28 to 37%)	29% (25 to 33%)	20% (16 to 24%)
4	1% (0 to 5%)	–	8% (5 to 12%)	4% (2 to 6%)	39% (34 to 44%)	29% (25 to 34%)	20% (17 to 24%)
5 (most deprived)	–	1% (0 to 3%)	9% (6 to 13%)	8% (5 to 11%)	47% (40 to 54%)	34% (29 to 40%)	21% (17 to 25%)
All	0% (0 to 1%)	0% (0 to 1%)	7% (6 to 9%)	4% (3 to 4%)	38% (36 to 40%)	26% (25 to 28%)	
p for trend	–	–	0.03	<0.001	0.04	<0.001	<0.001*

* Adjusted for age and sex.
Brackets contain 95% confidence interval.
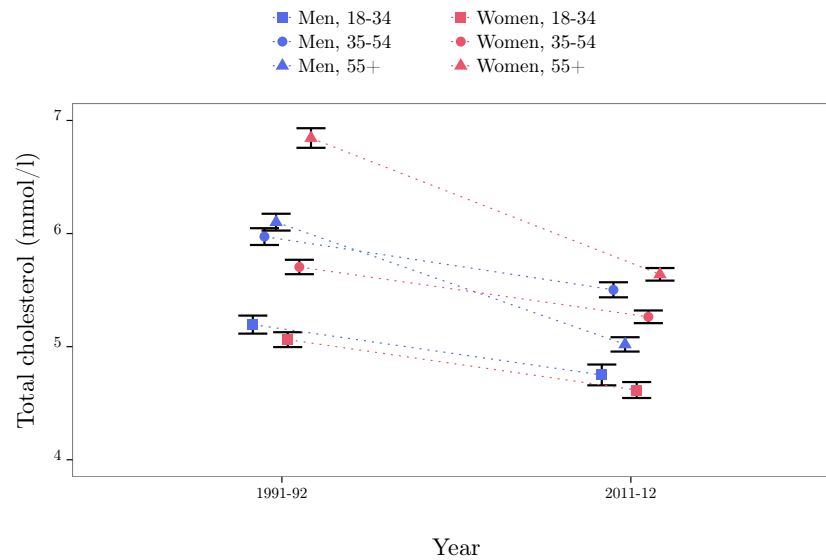
Figure 4.1: Observed mean total cholesterol (mmol/l) decline in England between 1991-92 and 2011-12 in men and women by age group. The error bars depict 95 % CI of the means. The vertical axis starts at 4 mmol/l to improve readability. The dotted lines are visual aids and do not reflect linear fits.

ment with the observed unit consumption of almost 4.07 billion, being just 1.5 % lower.[327, 328])

The mean total cholesterol of the adult community dwelling population in England decreased from 5.86 mmol/l (95 % CI: 5.82 mmol/l to 5.90 mmol/l) in 1991-92 to 5.17 mmol/l (95 % CI: 5.14 mmol/l to 5.20 mmol/l) in 2011-12. The decrease was observed in all age groups and it was steeper for ages over 55 for women and 35 for men (figure 4.1). The inverse socioeconomic gradient observed since 1998 persisted overall and in the subgroup of those aged over 55 years. No gradient was observed for other age groups (table 4.3 on page 111). On the contrary, I did not observe any socioeconomic gradient in 1991-92 with social class as a socioeconomic indicator when adjusted for age and sex (table 4.2 on page 110). The trend remained non-significant even when I placed the 'Other' social class group before all other groups.

4.3.1 *No statins scenario*

I estimated the total effect of statins on total cholesterol reduction using the equation (4.1) on page 106 as E_w = 25.7 % (95 % CI: 23.3 % to 28.0 %). The mean predicted total cholesterol TC_{pred} of the population was calculated to be 5.36 mmol/l (95 % CI: 5.33 mmol/l to 5.40 mmol/l).

Figure 4.2 on page 115 depicts the predicted mean total cholesterol of the population without the effect of statins, against the observed mean total cholesterol in 1991-92 and 2011-12, by age and sex. When the effect of statins was removed, the inverse socioeconomic gradient of cholesterol in the overall population disappeared (slope -0.01, 95 % CI:

Table 4.5: Estimated proportional contribution of statins to total cholesterol reduction since 1991-92, stratified by age group, sex and quintile groups of Index of Multiple Deprivation (QIMD).

QIMD	35 to 54 (years)		55 to 75		18 to 75	
QIMD	Men	Women	Men	Women	Men	Women
1 (least deprived)	14.0% (-19.2 to 41.9%)	4.2% (-24.4 to 28.3%)	50.6% (36.2 to 64.6%)	24.4% (11.4 to 36.6%)	33.5% (15.6 to 49.9%)	15.9% (1.9 to 28.9%)
2	13.0% (-28.2 to 45.9%)	5.9% (-23.9 to 30.7%)	59.7% (43.5 to 75.7%)	23.9% (10.2 to 36.8%)	36.0% (19.5 to 51.2%)	14.3% (2.0 to 25.5%)
3	17.9% (-49.7 to 78.2%)	3.6% (-33.9 to 33.3%)	37.8% (21.3 to 52.7%)	36.0% (21.2 to 50.1%)	26.9% (7.3 to 44.6%)	23.5% (7.5 to 37.5%)
4	29.0% (-7.2 to 58.9%)	19.3% (-37.9 to 64.4)	45.0% (27.7 to 60.6%)	36.3% (21.3 to 50.5%)	34.4% (19.0 to 48.9%)	24.8% (9.2 to 38.9%)
5 (most deprived)	31.1% (-9.1 to 63.8%)	37.1% (-19.0 to 79.9%)	43.2% (28.6 to 57.1%)	32.6% (16.0 to 47.9%)	33.8% (19.7 to 46.1%)	33.4% (18.3 to 47.5%)
All	22.2% (4.8 to 39.8%)	11.9% (-4 to 26.1%)	48.0% (40.1 to 56.1%)	40.0% (23.3 to 54.9%)	33.2% (25.8 to 40.6%)	21.3% (14.8 to 28.0%)
p for trend	0.24	0.03	0.41	0.17	0.99	0.02

Age group 18-34 was omitted, as statins' contribution was practically zero. Analysis was restricted to ages younger than 76 due to low number of older participants. Brackets contain 95% uncertainty interval estimated by Monte Carlo simulation.
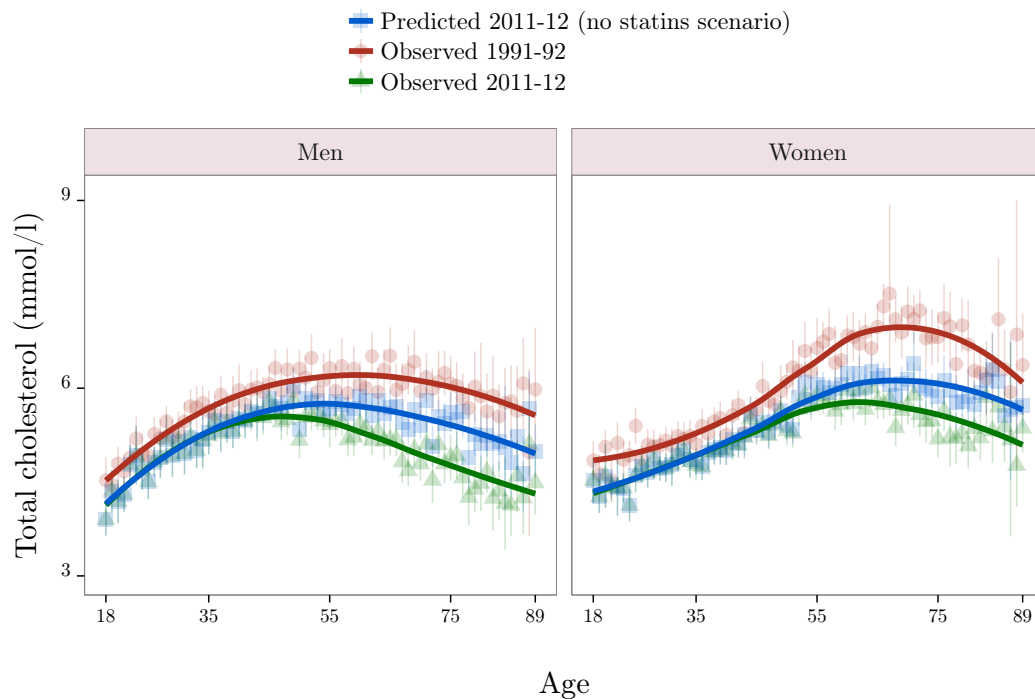
Figure 4.2: Mean total cholesterol by age, in men and women, in England (observed and predicted values). Points depict mean total cholesterol and error bars depict 95 % CI. The curves were derived from weighted local regressions. Due to small sample sizes I aggregated participants aged 89 with those older than 89 years. To improve readability the axes are not numbered from 0.

-0.03 to 0.01, p = 0.45). Subgroup analysis revealed that for men over 55 the slope was reduced to -0.05 (95 % CI: -0.10 to -0.01, p = 0.03) and for women over 55 the gradient was essentially zero (slope -0.04, 95 % CI: -0.08 to 0.01, p = 0.09). In addition, a socioeconomic trend appeared for women between 35 and 54 years with a slope of 0.05 (95 % CI: 0.01 to 0.10, p = 0.01). There was no other statistically significant gradient, for the remaining age groups.

Finally, statins were estimated as responsible for approximately 33.7 % (95 % UI: 28.9 % to 38.8 %) of the total cholesterol reduction since 1991-92. When stratified by sex statins contribution was 40.1 % (95 % UI: 33.6 % to 47.7 %) in men and 28.6 % (95 % UI: 22.3 % to 35.0 %) in women. Table 4.5 on page 114 summarises the contribution of statins for each socioeconomic group, by age group and sex. The negative values in the UI, implying that statins could have increased cholesterol to some, are an artefact of the Monte Carlo simulation due to wide mean cholesterol CI overlapping in some ages. The contribution of statins was consistently higher among men, consistent with the observed higher utilisation.

4.3.2 *Sensitivity analysis*

The mean predicted total cholesterol (TC_{pred}) of the population, using the inflated SE of E_w , was calculated to be 5.39 mmol/l (95 % CI: 5.35 mmol/l to 5.42 mmol/l). This is less than a 0.03 mmol/l difference from the main analysis. For the subgroup of deprived men older than 55, with the highest statin utilisation, the TC_{pred} from the sensitivity analysis was 0.09 mmol/l higher than the one from the main analysis. Similarly, the contribution of statins to the observed cholesterol decline for the whole population was estimated to be 33.9 % (95 % UI: 28.8 % to 38.7 %), a 0.2 % difference from the main analysis result. A similar pattern of minimal changes was observed for the remaining results.

4.4 DISCUSSION

This is the first study to quantify the contribution of statins to the observed decrease of total cholesterol in England's population by socioeconomic group. My results strongly suggest that the statins were not the main driver of total cholesterol reduction since 1991-92. In fact, only around one third of the overall reduction might be attributed to statins, and that was mainly in patients aged over 55 years and more so in men than women. Statins were more widely used in deprived than affluent areas. They appeared to help reduce socioeconomic inequalities in total cholesterol among women, but not among men.

4.4.1 *Utilisation of statins*

In this study, the utilisation of statins was higher in more deprived areas for men and women aged over 35 years. This socioeconomic pattern may partly reflect the higher prevalence of CVD in more deprived areas[334] and the incentivised use of the QRISK score for cardiovascular risk stratification in clinics, which includes area deprivation as a risk factor for CVD.[335, 336] My findings are consistent with earlier studies that used different methodologies. Ashworth *et al.* and Wu *et al.* also report that statin prescription was higher in more deprived areas in the UK.[337, 338] This success in tackling inequalities might be attributed to the NHS, since evidence from Australia, Sweden, Denmark and the US suggest that statin prescription in these countries has a socioeconomic gradient, with a less than expected utilisation among the more disadvantaged, and potentially increases health inequalities.[339–342]

4.4.2 *Contribution of statins to total cholesterol decline*

The second interesting finding is the contribution of statins to the observed decline of total cholesterol since 1991-92. I found that statins are not the main driver of the cholesterol decline in England, echoing studies from Iceland, Sweden, Finland and the Czech Republic.[309–312] I estimated that only about a third of the observed total cholesterol decline could be attributed to statins. This contribution was slightly higher than the aforemen-

tioned studies, perhaps reflecting a more recent time period with correspondingly higher statin use in England 2011-12, and possible nuanced differences in methodologies. While the cholesterol decrease was observed in all age groups since 1991, statins mostly contributed to the fall in people older than 55 years. Yet, some evidence exists which indicates that cholesterol changes in the fourth and fifth decade of life are important and influence CVD risk even 50 years later.[343]

On the contrary, using similar methodology to mine Bandosz *et al.* estimated that the contribution of statins to the observed decline of total cholesterol in Poland since 2002 was about 85 %.[344] The socioeconomic transformation of Poland in the 1990s led to significant improvements in diet and to a sharp decline in CVD mortality.[345] However, there have been no comprehensive population-wide policies targeting unhealthy diet in Poland since then. This is likely to be the explanation for the different results between England and Poland. Hence, this is a missed opportunity for Poland, rather than a case of a high-risk interventions being more effective than a population-wide one.

The observed inverse socioeconomic gradient in total cholesterol levels might be partly attributed to statins. In the no statins scenario the gradient disappeared completely when all ages were considered. However, the statin contribution varied across different genders and socioeconomic groups. Statin utilisation was higher in the most deprived groups, but inequitable by gender, reaching barely one third in women (34 %) but almost half (47 %) of deprived men in the 55+ age group. This difference can, at least partly, be explained by the higher CVD prevalence among men. By contrast, the statin contribution to cholesterol lowering was rather stable across socioeconomic groups in men (some 33 %), but rose from 16 % to 33 % in women. This suggests that the component of all other cholesterol reduction drivers had a higher impact among the most deprived men, while their effect among women of all socioeconomic backgrounds was more or less equal. This demands further research.

4.4.3 *Public health implications*

Overall, my research supports the principle of statins being the second best option for primary prevention. Non-statin interventions account for two thirds of the total cholesterol reduction observed since 1991-92, which can be attributed to dietary changes because physical activity levels have not increased substantially over this period.[308, 346] Indeed, United Nations Food and Agriculture Organisation data indicate that the animal fat supply per capita in the UK has fallen by almost 25 % since 1991.[347] This echoes Rose's original assertion that the greatest public health impact will be achieved through population-wide reductions in CVD risk than through interventions targeting high-risk individuals (section 1.5 on page 30).

Furthermore, the recent proposed widening of criteria for statin prescription in primary prevention by the American College of Cardiology and the American Heart Association[315] and NICE[348] has been questioned on grounds of effectiveness, cost-effectiveness, acceptability and safety.[318] These measures may prove to be less effective than anticipated because of cumulative attrition factors. Approximately half of the UK patients that are com-

menced on lipid lowering medication for primary prevention are ineligible according to the respective guidelines, while many eligible patients remain untreated.[338] Moreover, over half the patients commenced on statins for primary prevention have discontinued them within two years.[349–352] In addition to medicalising otherwise healthy individuals, some patients may also be tempted to adopt more unhealthy diets because of the false reassurance that statins will compensate for the unhealthy behaviours.[353] Along with the increased resource requirements, an additional opportunity cost comes from undermining the primary driver of cholesterol decline – nutritional improvements at individual and national policy levels.[354]

Regarding inequalities in healthcare, my research suggests that English statin prescribing might be equitable. These results are intriguing, because healthcare based interventions generally are expected to increase the inequality gap (section 1.5 on page 30). This highlights why policy-making cannot be based only on theoretical evidence about the equity of a proposed policy, and why modelling is vital in equity assessment. In this particular case, there are at least three factors that are likely to have improved the equity of statin utilisation. First, this represents a success for the socialised medicine provided by NHS England. Second, clinicians in England use the QRISK risk assessment algorithm to estimate the risk of individuals of developing CVD in the next decade. QRISK includes deprivation as one of the risk factors for CVD; therefore, it incentivises statin prescription for the most deprived. Finally, CVD prevalence increases with deprivation; hence, the proportion of statin utilisation that reflects secondary prevention increases with deprivation.

4.4.4   *Strengths and limitations*

This study was grounded on the best available evidence to explore the research question. I integrated all the available data from HSE, a high quality cross-sectional survey, the Prescription Cost Analysis report, an accurate and precise report about prescriptions in England, and published meta-analyses on the effect of statins. The modelling approach allowed for the best use of all the available information. In fact, despite the assumptions regarding the effects of statins my results were robust to the sensitivity analysis. Any biases and errors were diluted because they only applied to about 13 % of the sample who were statin users.

However, this study has several limitations. First, it is based on self-reported statin prescription and adherence; nevertheless, consistent data from prescription cost analysis reports for 2011-12 suggest that the estimated prevalence of statin utilisation is fairly accurate. Second, unlike HSE2011-12, HSE1991-92 was not weighted to adjust for non-response bias. Furthermore, no other HSE has recorded statin use separately from other lipid lowering medication; this renders an interim point analysis between 1991 and 2011 practically impossible. Third, there were no common or directly compatible socioeconomic indicators between the two surveys to allow for more accurate comparisons. My assumption that there was no socioeconomic gradient of mean total cholesterol in 1991-92 is suggested by my finding of no such gradient by social class in HSE1991-92. This is consistent with Scholes *et al.* who also found no socioeconomic gradient in 1994 using QIMD as socioeco-

nomic indicator.[313] The Whitehall II cohort also showed no socioeconomic gradient for total cholesterol in 1985-88.[355] Neither did my analysis consider other inequalities, for instance, ethnic minorities or people with mental health or illiteracy problems.[337, 356, 357] Fourth, the estimate of the statin effect E_w was derived mostly from short term trials lasting less than one year. Edward *et al.* have shown that the statins effect remains fairly stable in trials lasting more than one year.[325, Additional file 5] In addition, the estimation of E_w assumes that the differences between each trial population and my study sub sample of statin users were the same for each statin. Finally, I defined statin users as those who had taken a statin at least once during the week before the nurse interview. Since adherence is much higher in trials, my statins effect is probably an overestimation.[358]

4.5 CONCLUSIONS

This chapter suggests that statins contributed about one third of the observed total cholesterol decline in England since 1991-92. Their impact on reducing socioeconomic inequalities in total cholesterol was generally positive, even though it being an agentic intervention. In England, statin prescription for primary CVD prevention is mainly promoted through universal screening of the population; a programme known as NHS Health Checks. In the next chapter I will explore the effectiveness and equity of NHS Health Checks.

5

CARDIOVASCULAR SCREENING FOR PRIMARY PREVENTION

5.1 INTRODUCTION

The unquestionable efficacy of pharmacological treatments for secondary prevention of CVD primed the notion that similar medication would also be effective for primary prevention. If treatment can substantially reduce the risk of a new CVD event after one has already occurred, why not use the same treatment options on healthy individuals to prevent the first CVD event from ever occurring? This, perhaps oversimplified, argument seems a reasonable and particularly attractive option; especially among clinicians and policy makers. The updated guidance to widen the criteria for statin prescription in primary prevention is a recent example (chapter 4 on page 103). Yet, this approach may pose risks and inefficiencies when real life implementation is considered.

In England, the current governmental action plan to tackle CVD burden includes a programme known as NHS Health Checks. Introduced in 2009, this programme promotes the screening of all healthy individuals aged 40 to 74 for CVD risk stratification and treatment of those at high-risk,[124, 359] an approach similar to the simplified argument of the previous paragraph. However, the debate regarding the scientific foundation, effectiveness and cost-effectiveness of this approach has been recently heated.[360–364] Despite this controversy, the programme remains policy.

Beyond the obvious importance of the debate to national public health, its relevance extends internationally. Public health policy choices in the UK influence policy world wide; the UK tobacco control and salt reduction strategies are two recent examples.[365, 366] In essence, the debate about NHS Health Checks originates from the archetypal debate of targeted 'high-risk' versus 'population-wide' preventive interventions that was first articulated by Geoffrey Rose (section 1.5 on page 30). In Rose terminology, NHS Health Checks is the typical 'high-risk' intervention, since it targets high-risk individuals, rather than lowering risk in the whole population.

The effectiveness of high-risk interventions for CVD prevention has been challenged before.[143] More recently, a Cochrane systematic review and the Inter99 trial have found no benefits of health checks on CVD morbidity or mortality.[367, 368] There were, however, major limitations to these studies: Inter99 trialled a counselling intervention not supported by additional pharmacological treatment; and in the Cochrane review 9 out of 14 trials were conducted before 1980, when the treatment options for high-risk individuals were very limited. In addition, high-risk interventions may be more effective in populations with high clustering of risk factors, resulting in high concentration of the risk to certain groups in the population.[369] In fact, the English population has such characteristics, with the concentration of CVD risk being higher among the most socioeconomically deprived groups (section 1.4 on page 23).[313]

High-risk interventions may generate health inequalities because they require active participation of individuals in both screening and treatment of those at high-risk, favouring those with more resources (please refer to primary prevention typologies, section 1.5 on page 30). The particular effect of NHS Health Checks on socioeconomic health inequalities however, remains unclear. Two national studies reported no difference in the coverage of the intervention by deprivation and a slightly higher attendance among those living in the most deprived areas;[370, 371] however, the actual uptake[38] was not studied. In contrast, several smaller but more detailed studies showed significantly lower uptake in deprived areas.[372–374]

The aim of this chapter is to estimate the potential impact of universal screening for primary prevention of CVD on disease burden and socioeconomic health inequalities in England. Available data on the effectiveness of the NHS Health Check programme have been used to model this scenario. I further compared universal CVD screening with: 1. an alternative approach targeting only deprived areas; 2. with a feasible population-wide intervention; and 3. with a combination of 1 and 2.

5.2   METHODS

This is the first results chapter that uses the dynamic features of the IMPACT$_{NCD}$ model as have been described in chapter 2 on page 41.[39] The projection horizon was set at 2030 for this simulation, and the following scenarios were simulated.

5.2.1   *Scenarios*

BASELINE (CURRENT TRENDS):   In the baseline scenario, I assumed that the recent observed trends in CVD risk factor trajectories by age, sex, and socioeconomic status will continue in the near future. This is a similar scenario to the one that was described in the Methods chapter (chapter 2 on page 41) and it was used for the validation of the model (chapter 3 on page 77).

UNIVERSAL SCREENING: This scenario modelled the potential health effects of universal screening to identify and treat people at high-risk for CVD. Input variables were informed from current implementation of the NHS Health Check programme. Eligible people were defined as adults aged between 40 and 74, excluding those with a known history of CVD, atrial fibrillation, diabetes mellitus, rheumatoid arthritis, or renal disease; closely resembling real life eligibility criteria. Based on existing evidence, I assumed an

---

38 Uptake is calculated by dividing those who attended by those who have been invited. It differs from coverage, which has all eligible population in the denominator. This is important because if some practices invited first those living in deprived areas, coverage does not reflect the true equity of the intervention.

39 More accurately, this chapter represents work that I did during the first and second year of my PhD. Smoothing of the model inputs had not been developed at that time, HSE2011 was used to prime the synthetic population and the lag time was deterministic. The technical specification of the model version that was used for this chapter has been published and is available at http://www.bmj.com/highwire/filestream/924761/field_highw ire_adjunct_files/0/kypc031638.ww1_default.pdf.

uptake of 50 % for screening,[375] and to mirror the UK population, I calibrated the distribution of the estimated 10 year risk of developing CVD among those participating: 70 % with a less than 10 % risk, 25 % with between 10 % and 20 %, and 5 % with more than 20 %.[370] In addition, I calibrated the age distribution so that around 30 % of those screened were older than 60.[370] Participants with a higher than 10 % estimated 10 year risk of developing CVD were considered at high-risk and eligible for treatment. I used the QRISK2 score to estimate the perceived from healthcare 10 year risk of developing CVD.[336]

Based on published evidence, I assumed that about 24 % with an estimated risk of 20 % or more and total cholesterol of 5 mmol/l or more will be prescribed Atorvastatin 20 mg and about 27 % with an estimated risk of 20 % or more and a SBP of 135 mmHg or more will be prescribed antihypertensive medication. For those with a risk between 10 % and 20 % I assumed that about 17 % and 20 % will be prescribed treatment, respectively.[376] I assumed an 80 % overall persistence to continue prescribing the medication and a mean adherence of approximately 70 %, roughly based on evidence from Denmark.[352] Moreover, I modelled high-risk participants with a body mass index of more than 50 kg/m² to undergo bariatric surgery and reduce their BMI to 30 kg/m². I assumed that with lifestyle counselling half of the high-risk participants consuming fewer than five fruit and vegetable portions daily will increase their consumption by a portion daily. Half of those being active for less than five days a week will increase their physical activity by an active day each week, and all high-risk participants will decrease their BMI by around 1 %.[376, 377] Finally, I modelled 10 % of high-risk smokers to achieve cessation for a year and have a probability of relapse equal to that of the general population by sex, QIMD, and years since cessation.[378, 379]

CONCENTRATED SCREENING: In the concentrated screening scenario, I simulated a hypothetical strategy where screening had only been implemented in the two most deprived fifths (QIMD 4 and 5), the groups with the greatest concentration of CVD risk. I assumed that the uptake of the intervention was 50 % and the risk and age distribution in the participants was similar to that in the eligible population. Otherwise, the strategy is similar to the previous universal screening scenario. Given the recent criticism about the cost and cost-effectiveness of the intervention,[363] offering the intervention where the risk is more concentrated may reduce costs.

POPULATION-WIDE INTERVENTION: This scenario modelled the effects of a feasible population-wide structural intervention targeting unhealthy diet and smoking. Several studies have found that a tax on sugar sweetened beverages may reduce the prevalence of obesity.[380–382] For this scenario I assumed that such a tax may reduce the mean increase in BMI by about 5 % annually. Moreover, the UK has had one of the world's most successful salt reduction strategies, including public awareness campaigns, food labelling, and voluntary reformulation of processed foods.[383] Modelling studies suggested that the addition of mandatory reformulation of processed foods may further reduce mean SBP by 0.8 mmHg,[176] I modelled this decrease. A large randomised trial in the US showed that subsidies on fruits and vegetables may increase consumption by about half a portion daily and a modelling study in the UK found that subsidising fruits and vegetables combined

with taxation of unhealthy foods may increase fruit and vegetable annual consumption by about 10 %.[384, 385] I modelled an increase of a portion of fruit and vegetable each day in 50 % of the population. Finally, a SimSmoke modelling study estimated that full compliance with the framework convention on tobacco control might reduce smoking prevalence by 13 % (relative) in five years;[173] I modelled this decrease.

THE COMBINATION OF POPULATION-WIDE INTERVENTION AND CONCENTRATED SCREENING: This scenario is the combination of the population-wide intervention and concentrated screening strategies. I modelled the implementation of a population-wide strategy identical to the previous scenario, complemented by concentrated screening for people at high-risk of CVD in the most deprived fifths (QIMD 4 and 5).

5.2.1.1 *Common scenario assumptions*

All interventions began in 2011 and were linearly diffused into the population over a 5-year period. Trends in population risk factors were assumed to be the same as those of the baseline scenario for all but the population-wide intervention. All of the scenarios assumed that CVD case fatality will keep improving by 3 % (relative) annually. In addition, I assumed a socioeconomic gradient in CVD case fatality, forcing the more deprived people to experience worse outcomes. Both case fatality assumptions were based on recent trends and are suggested by the British Heart Foundation's statistics on coronary heart disease.[92] Finally, a 5-year lag time was assumed between exposure to cardiovascular risk factors and disease. A more detailed scenario specification can be found in the appendix (appendix C.3 on page 246).

5.2.1.2 *Model outputs*

I report the cumulative estimates of cases and deaths prevented or postponed as measures of overall effectiveness of the modelled interventions. To measure the impact of the modelled interventions on absolute and relative socioeconomic health inequalities, I used two regression based metrics inspired by the slope index of inequality: the absolute equity slope index; and the relative equity slope index (both described in section 2.7.2 on page 64). The absolute equity slope index measures the impact of an intervention on absolute inequality; for example, a value of 100 means 100 more cases were prevented or postponed in most deprived areas compared with least deprived areas, resulting in a decrease in absolute inequality. The relative equity slope index takes into account the pre-existing socioeconomic gradient of disease burden and measures the impact of an intervention on relative inequality. Positive values mean the intervention tackles relative inequalities and negative values that the intervention generates relative inequality. Finally, I summarised the overall impact of each scenario on CVD burden and equity in the equity summary chart.

5.2.1.3 *Uncertainty and sensitivity analysis*

As I have described in the Methods chapter (section 2.6 on page 62), IMPACT$_{NCD}$ implements a second order Monte Carlo design that allows uncertainty to be quantified from the outputs. The probabilistic sensitivity analysis has been incorporated in my reported estimates. I summarise output distributions by reporting medians and interquartile ranges in the form of first and third fourths. Table A.1 on page 187 provides a detailed description of the relevant distributions that have been used as inputs, and their sources.

I ran three further scenarios offering slight variations on the two primary ones of universal screening and population-wide intervention: a universal screening variation, where I assumed a treatment threshold recommendation of 20 % risk instead of 10 %; another variation on universal screening, where I assumed a socioeconomic differential in screening uptake, with the most deprived of the population to be 10 % less likely to participate; and a variation on the population-wide intervention, where I only modelled dietary interventions, excluding smoking interventions.

5.2.1.4 *Validation*

For the validation of this version of the model I followed the same approach that I described in chapter 3 on page 77.

5.3 RESULTS

IMPACT$_{NCD}$ outputs for burden and inequality are summarised for ages 30 to 84. Because of the assumed 5-year lag time, the interventions affect the population from 2016 up to the projection horizon of 2030.

5.3.1 *Overall effectiveness*

Under the baseline scenario, IMPACT$_{NCD}$ estimated about 1.4 million (IQR: 1.3 to 1.5) cases of CVD and 540 000 CVD deaths (IQR: 520 000 to 550 000) between 2016 and 2030. The most effective intervention was the combination of the population-wide intervention and concentrated screening. The population-wide intervention alone had the second highest effectiveness, whereas the universal and the concentrated screening scenarios were considerably less effective (table 5.1 on the following page). Despite the improvement of most related risk factors, the proportion of high-risk people in the eligible population is slowly increasing over time, because of population ageing (figure 5.1 on the next page).

5.3.2 *Socioeconomic inequalities*

When socioeconomic inequalities were considered, the patterns for reductions in absolute and relative inequalities were similar. The combination of the population-wide intervention and concentrated screening seemed the most powerful among the simulated interven-

Table 5.1: Estimated CVD cases and deaths prevented or postponed under each scenario, by 2030.

Scenarios	Number (interquartile range) prevented or postponed	
	Cases	Deaths
Universal screening	19 000 (11 000 to 28 000)	3000 (−1000 to 6000)
Concentrated screening	17 000 (9000 to 26 000)	2000 (−100 to 5000)
Population-wide intervention	67 000 (57 000 to 77 000)	8000 (4000 to 11 000)
Population-wide intervention and concentrated screening	82 000 (73 000 to 93 000)	9000 (6000 to 13 000)

Results rounded to nearest 1000.
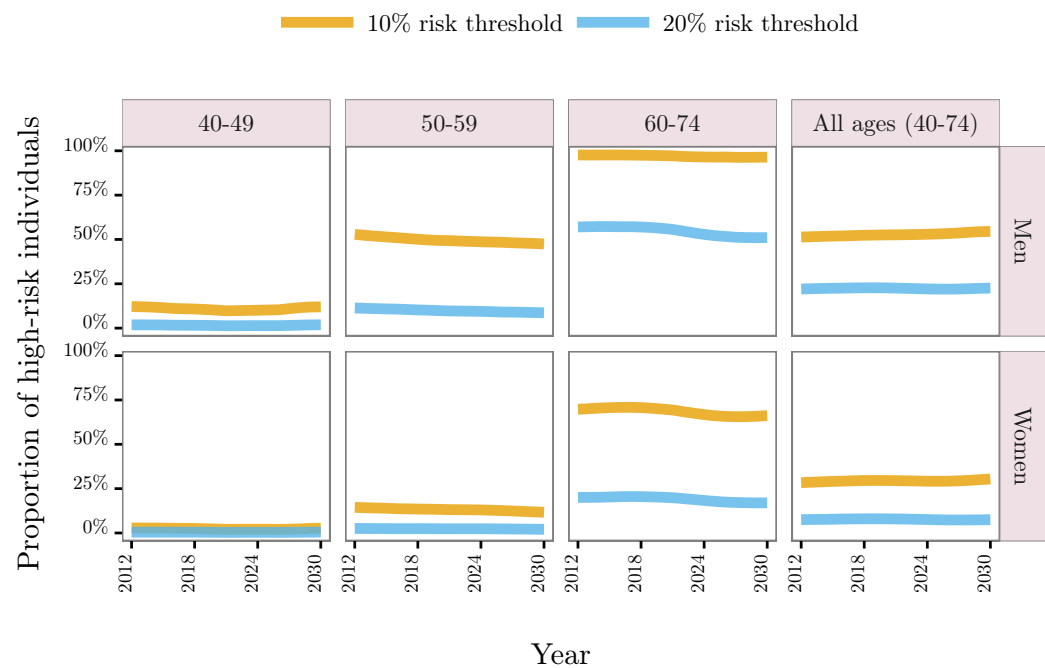
Figure 5.1: Proportion of high-risk people eligible for universal screening. Population projections, by age group and sex. 10 year risk of CVD was estimated from QRISK2 score.

tions (table 5.2 on the following page and table 5.3 on page 129). Concentrated screening alone was the second most powerful intervention in tackling inequalities, followed by the population-wide intervention. Finally, universal screening of CVD is likely to have a small or negligible effect on socioeconomic inequalities.

5.3.3 *Equity summary chart*

I summarised the estimates for the effectiveness and equity of the modelled interventions in the equity summary chart (figure 5.2 on page 130). The horizontal axis of the chart represents the cases of CVD prevented or postponed and the vertical axis the reduction in absolute inequality. Scenarios above the equity curve (represented by the dashed curve in the figure) decrease relative socioeconomic inequality, and scenarios below the curve increase it. The vertical distance from the curve approximates the impact of the scenario on relative inequality (please refer to section 2.7.2.2 on page 66). The combination of the population-wide intervention and concentrated screening is by far the most effective and equitable intervention. Concentrated screening is also equitable but with fewer morbidity gains.

5.3.4 *Sensitivity analysis*

Adding assumptions to extend the scenarios did not displace my main findings. The three most notable results of the sensitivity analysis were: 1. raising the treatment threshold from 10 % to 20 % further reduced the effectiveness of universal screening by about 60 % in preventing CVD cases. However, in preventing deaths from CVD the effectiveness decreased by only 15 % as raising the treatment threshold excludes younger participants at intermediate risk from treatment. 2. Assuming a differential uptake of universal screening by deprivation fifth essentially eliminated the estimated small potential benefit of universal screening in tackling health inequalities. 3. A population-wide intervention targeting only diet would still be about twice as effective as universal screening and more than twice as effective as population-wide intervention targeting smoking alone – so the relative ranking of scenario effectiveness would remain unaltered. Appendix C.4 on page 248 provides detailed model outputs for the extra scenarios.

5.3.5 *Validation*

The full validation of this version of IMPACT$_{\text{NCD}}$ has been published and is available at http://www.bmj.com/highwire/filestream/924761/field_highwire_adjunct_files/0/kypc031 638.ww1_default.pdf. The validation is very similar to the validation of the latest version of the model that has been presented in chapter 3 on page 77. Figure 5.3 on page 131 summarises the validation process.

Table 5.2: CVD cases prevented or postponed according to fifth of deprivation by 2030, along with absolute equity slope index for each scenario.

Deprivation fifth*	Number (interquartile range) of cases prevented or postponed			
	Universal screening	Concentrated screening	Population-wide intervention	Population-wide intervention and concentrated screening
1 (least deprived)	3400 (−1400 to 8300)	0	10 800 (5900 to 15 500)	10 800 (6200 to 15 700)
2	2900 (−1500 to 8400)	0	12 200 (6200 to 17 200)	11 500 (6600 to 17 000)
3	4000 (−900 to 9300)	0	13 100 (8100 to 18 300)	12 600 (7400 to 17 700)
4	3700 (−1600 to 8600)	6400 (1500 to 11 800)	12 500 (7100 to 18 400)	18 700 (13 900 to 24 200)
5 (most deprived)	4900 (−600 to 10 400)	10 700 (5300 to 16 300)	18 700 (13 000 to 24 000)	28 600 (22 800 to 33 200)
Absolute equity slope index	1700 (−6200 to 9300)	14 100 (5700 to 23 000)	8400 (−400 to 16 900)	21 100 (12 800 to 29 300)

*According to Index of Multiple Deprivation.
Results rounded to nearest 100.

Table 5.3: Relative percentage reduction in cases of CVD according to fifth of deprivation by 2030, along with relative equity slope index for each scenario.

Deprivation fifth*	Relative percentage reduction (interquartile range)			
	Universal screening	Concentrated screening	Population-wide intervention	Population-wide intervention and concentrated screening
1 (least deprived)	1.3 (−0.5 to 3.1)	0	4.1 (2.2 to 5.9)	4.0 (2.4 to 6.0)
2	1.1 (−0.5 to 2.9)	0	4.2 (2.2 to 5.9)	4.0 (2.3 to 5.9)
3	1.4 (−0.3 to 3.2)	0	4.6 (2.8 to 6.3)	4.4 (2.6 to 6.2)
4	1.3 (−0.6 to 3.1)	2.4 (0.6 to 4.3)	4.6 (2.7 to 6.6)	6.9 (5.1 to 8.9)
5 (most deprived)	1.6 (−0.2 to 3.3)	3.6 (1.8 to 5.3)	6.2 (4.4 to 8.0)	9.4 (7.6 to 11.2)
Relative equity slope index	0.4 (−2.4 to 3.2)	4.9 (1.8 to 7.9)	2.3 (−0.7 to 5.3)	6.7 (3.8 to 9.5)

*According to Index of Multiple Deprivation.
Results rounded to one decimal place.
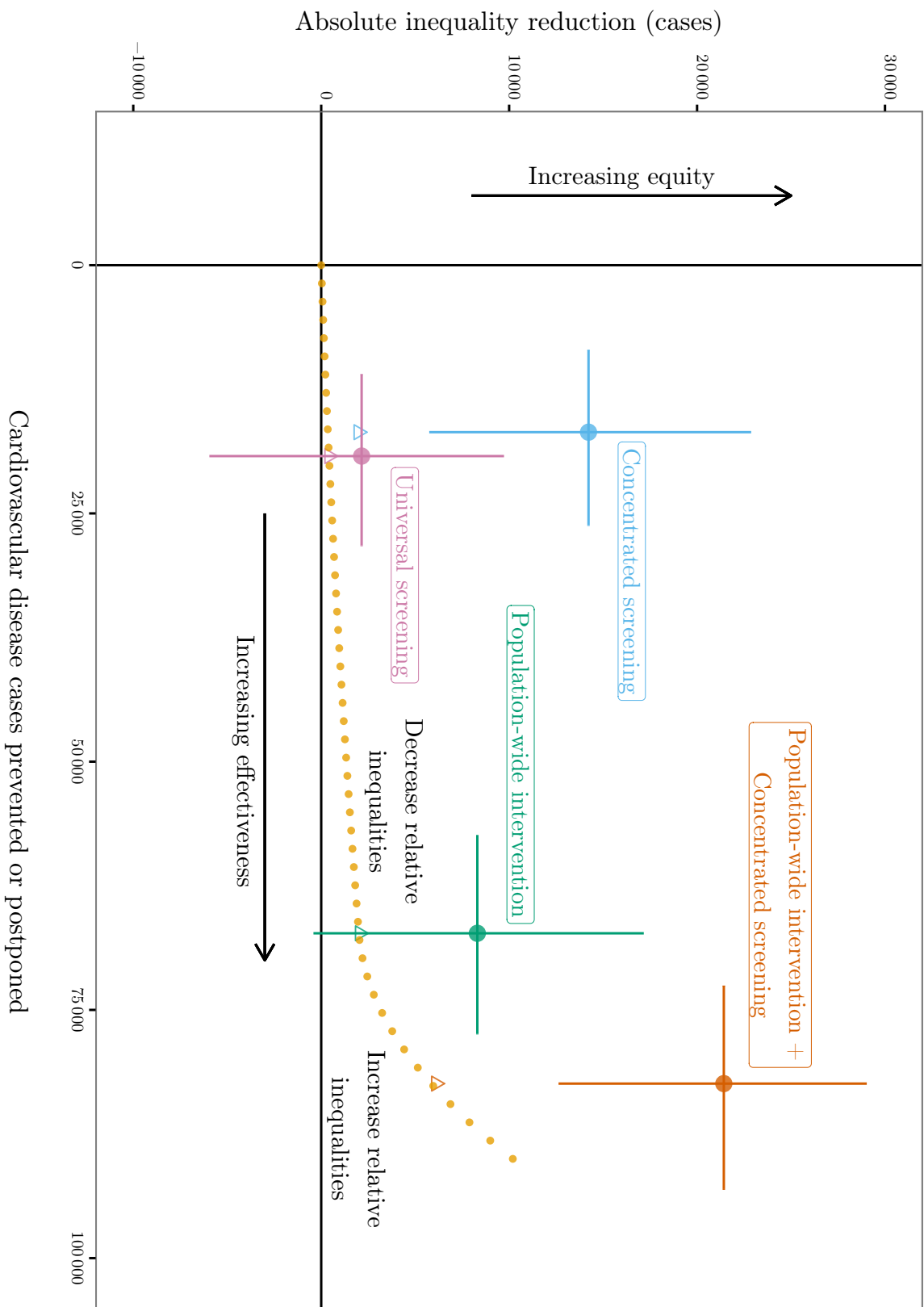
Figure 5.2: Equity summary chart of effectiveness and equity of all modelled interventions, compared with baseline scenario (beginning of axes). The dotted curve represents the 'equity' curve. Interventions below the curve increase relative inequality, whereas interventions above it decrease relative inequalities. Smaller coloured triangles represent reference points used to fit equity curve. Horizontal and vertical error bars represent interquartile ranges.
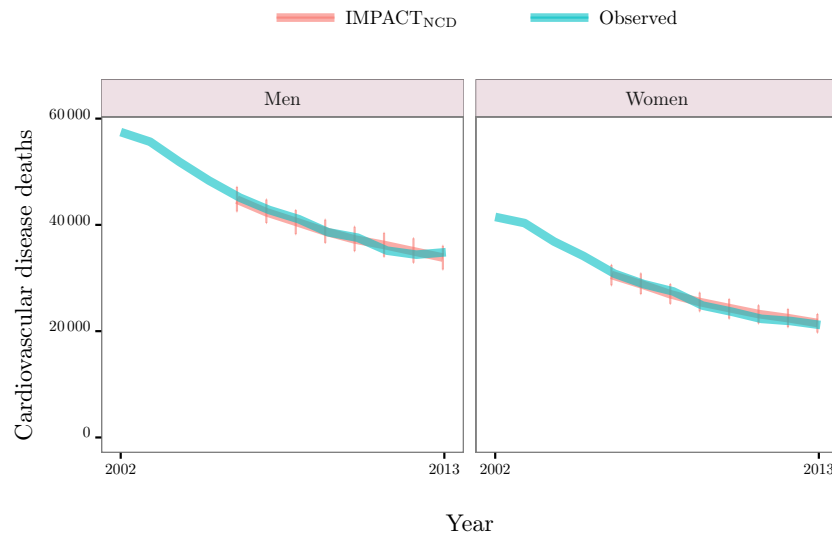
Figure 5.3: Number of deaths from cardiovascular disease (CVD) in England, by year for ages 30 to 84. Office for National Statistics reported deaths (observed) versus IMPACT$_{NCD}$ estimated. Observed deaths after 2010 were adjusted to account for changes in ICD10 version used by the Office for National Statistics from 2011 onwards. Error bars represent interquartile ranges.

5.4 DISCUSSION

My results strongly suggest that universal screening and treatment of people at high-risk is not the most effective option for primary prevention of CVD overall, nor for reducing socioeconomic inequalities. In contrast, prevention strategies that include population-wide structural interventions seem to be the consistently better options for reducing overall CVD burden and inequalities. This echoes and quantifies findings from other, mostly theoretical, studies suggesting that structural population-wide interventions are powerful, while reducing socioeconomic health inequalities.[112, 129, 140, 143] Indeed, the impact of the population-wide intervention scenario on reduction in estimated mortality and inequalities seems compatible with previous estimates, considering the different methodologies.[175] Furthermore, the effectiveness and equity of population-wide structural interventions can be further improved by the addition of targeted interventions in the most deprived groups, as highlighted in the combined scenario of the population-wide intervention and concentrated screening.

Compared with other modelling approaches, my IMPACT$_{NCD}$ model estimated that NHS Health Checks might prevent approximately 1000 non-fatal and 200 fatal cases of CVD annually. This is comparable with the Department of Health estimates of 1600 non-fatal CVD cases and 650 deaths prevented annually.[124] The difference in two estimates may be explained by the Department of Health modelling approach to assumed an intervention uptake of 75 %, higher than the current observed levels, and to use a static baseline from 2006 for CVD cases ignoring the downward trends.[386] Using the Archimedes model,

Schuetz *et al.* estimated that health checks in the UK could prevent some 12 CVD cases per 1000 population screened after 30 years follow-up (7500 CVD cases prevented each year extrapolating to the eligible English population).[387] That higher estimate reflects the researchers' apparently unrealistic assumption of 100 % screening uptake and 50 % overall uptake of treatment.

5.4.1   *The scenarios*

I modelled the universal screening scenario to closely resemble the current implementation of the NHS Health Check programme, based on published evidence. When the evidence were inconclusive, I assumed optimal implementation of the policy.[40] Therefore, I maintain that my estimates on the effectiveness of this scenario are not far, and perhaps overestimate, the real world effectiveness of NHS Health Checks. In addition to modelling assumptions, the estimates in my thesis reflect conventional assumptions of statin effect sizes; these benefits may have been overestimated according to some authors.[388] Moreover, my output suggesting that universal screening might reduce socioeconomic inequalities seems to contradict existing empirical and modelling evidence.[143, 144, 389, 390] This is because I generously assumed identical screening uptake and treatment adherence for all socioeconomic groups. In fact, any potential reduction in socioeconomic health inequalities was essentially eliminated when I considered a small socioeconomic differential in uptake in the sensitivity analysis. Furthermore, additional health inequalities may arise from differential persistence and adherence to treatment by deprivation status.[352]

   The population-wide intervention scenario on the other hand, is based mostly on structural policies targeting price and availability. This scenario's potential effectiveness was mostly based on natural experiments,[345, 391] and on previous modelling studies from the UK and elsewhere. The size of the changes in the population risk factors that I modelled were modest, and actually smaller than the reductions observed in countries such as France, Finland, and the US during recent decades.[305, 392, 393] This scenario estimated the reduction in mortality conservatively, because it ignored the beneficial effect of the policies on survival from CVD. Similarly, it underestimated the reduction of the gap in inequalities, because it did not fully consider the current disproportionate burden of poor diet among the most deprived of the population, and hence the potential for improvement through population-wide policies.[88, 89]

   Finally, the concentrated screening strategy was the weakest in terms of overall effectiveness, yet more powerful in tackling inequalities. Its increased impact on socioeconomic health inequalities is a direct consequence of the concentrated prevention only to the more deprived quantiles of the population. However, the scenario assumptions may not fully

---

40 This occurred in two occasions. The first was the assumption that universal screening has no differential uptake by deprivation. I tested this assumption in the sensitivity analysis. The second occasion was the assumptions regarding lifestyle improvements for screening participants. I did not include this assumption in the sensitivity analysis because universal screening scenario was less effective than alternative scenarios, despite the likely overestimation of their effectiveness.

hold in real world implementation. Hence, concentrated screening represents a challenge for public health practitioners and policy makers to exploit the opportunity of a smaller and more homogeneous eligible population and to implement better recruitment and tactics for treatment adherence. Yet, cost-effectiveness might also fall because of loss of economies of scale.

5.4.2 *Public health implications*

This IMPACT$_{NCD}$ modelling may help stakeholders to understand better the interplay between preventive policies, risk factors, disease, and inequalities; thus, potentially inform health policy and strategy. Hence, when compared with the alternative feasible interventions, universal screening seemed inferior both in primary prevention and in reducing socioeconomic health inequalities. Additionally, I estimated that the proportion of young people at high-risk aged less than 60 in the eligible population will decrease in future (figure 5.1 on page 126). This will render universal screening less effective and less cost-effective for this age group, because a larger number will need to be screened to identify each high-risk individual.

This study suggests that despite the high clustering of risk factors in the most deprived parts of the population, structural population-wide approaches remain more effective than high-risk ones for the prevention of CVD. Population-wide approaches also seem to be more effective in reducing absolute and relative socioeconomic health inequalities, generally cost much less than a universal screening programme, and may even be cost saving.[131, 394] In this study, I did not model the full potential of these policies, as I focused only on diet and smoking interventions; I did not, for example, incorporate alcohol consumption or physical activity. In addition, I did not simulate the likely wider benefits of improved diet and smoking cessation on the plethora of relevant NCDs. Despite this restricted scope, for CVD prevention I estimated that structural policies targeting diet could be twice as effective as those targeting smoking. Yet, structural interventions for a healthier diet are currently underutilised compared with tobacco control. Several countries have now introduced taxes on sugary drinks or sugar, including Finland, France, Latvia, and Mexico. The UK has recently followed their example. Hungary is the only European country currently taxing unhealthy 'junk' food.[395] However, fiscal interventions may face opposition from commercial vested interests.[396] Interestingly, an increasing body of evidence from empirical studies and modelling analyses suggest that the maximum health impact with a neutral effect on poverty may occur when food or drinks taxes are combined with subsidies for healthy foods.[385, 397, 398]

Moreover, the combination of a population-wide intervention with an intervention targeting the most deprived members, may further improve effectiveness and equity. This approach is in the spirit of proportionate universalism that was identified in the Marmot review as the best approach to tackle socioeconomic inequalities in health (please refer to section 1.5.3 on page 32).[98] This study provides evidence that in CVD prevention proportionate universalism may be the best option not only for tackling inequalities but also

for overall effectiveness. Moreover, it shows that targeted interventions can be effective despite the expressed opposite arguments.[98, 150]

5.4.3   *Strengths and limitations of this study*

IMPACT$_{NCD}$ is the first microsimulation model to synthesise core principles of social and CVD epidemiology, vital demographics, published literature, and recent health surveys for England to create a synthetic population of England, including socioeconomic structure, at the individual level. The microsimulation approach allows for the simulation of detailed scenarios and explores the distributional nature of their impact on the population, in a competing risks framework. Microsimulation allows for greater flexibility and more detailed simulation, demanding more statistical and computational resources than older approaches. Many assumptions must be made with such models. Yet, despite the potential frailty of such assumptions, this model validated well against observed CVD mortality, even when multiply stratified.

Models are simplifications of reality and thus possess inherent limitations. At least four items were not included in the current model. Firstly, the multiplicative risk assumption is considered the status quo in comparative risk assessments;[30] however, this may oversimplify the complex nature of interactions between multiple risk factors and disease outcome over the life course. Secondly, IMPACT$_{NCD}$ currently ignores the effect of risk factors on CVD case fatality, although in this study I considered only primary prevention scenarios. Thirdly, complex population dynamics such as migration, social mobility, and the socioeconomic consequences of disease were not modelled. I consider this bias would be relatively small for projections with a short horizon. Fourthly, the model ignores the impact of universal screening in recognising previously undiagnosed cases of atrial fibrillation and other opportunistic diagnoses. Reassuringly, most of these biases apply across all scenarios; their effects would thus be reduced in comparisons between scenarios.

5.5   CONCLUSIONS

When comparing primary prevention strategies for reducing CVD burden and inequalities, universal screening seems less effective than alternative strategies that incorporate population-wide approaches. Further research is needed to identify the best mix of population-wide and risk targeted CVD strategies to maximise cost-effectiveness and minimise inequalities.

In this chapter I mentioned that England has one of the most comprehensive strategies to reduce salt consumption, worldwide. I also modelled an increment of this step in the context of the population-wide intervention. In the next chapter I will specifically focus on existing and proposed population-wide policies to reduce salt consumption. Apart from the obvious importance from a public health perspective, the next chapter is important from a technical point of view because I will use modelling to simulate CVD and gastric cancer simultaneously.

SALT REDUCTION POLICIES

6.1 INTRODUCTION

In the previous chapter I modelled existing high-risk preventive policies and compared them with structural population-wide alternatives. In this chapter I will model an existing population-wide prevention policy that targets excess salt consumption in the population. There are two interesting elements in this policy. First, in the agentic – structural continuum (section 1.5 on page 30) this policy 'sits' somewhere in-between the two extremes; therefore, the equity of the policy cannot be easily assessed even in qualitative terms. Second, excess salt is a risk factor both for CVD and gastric cancer; therefore, this policy is a great example of NCD joint prevention, and it highlights the ability of IMPACT$_{NCD}$ to model separate lag times for each disease.

Excess salt consumption is associated with higher risk of CVD and gastric cancer.[56, 57] Globally, more than 1.5 million CVD related deaths every year can be attributed to excess salt intake.[59] Further salt related deaths come from gastric cancer. Health policies worldwide therefore aim to reduce dietary salt intake.[399] Furthermore, the WHO recommends reducing population exposure to salt as one of the 'best buy' strategies to prevent NCDs, highlighting its cost-effectiveness and feasibility.[29]

Since 2003, the UK has had one of the world's most successful salt reduction strategies, including public awareness campaigns, food labelling, and 'voluntary' reformulation of processed foods.[383] This package of measures is regularly evaluated and has been monitored through nationally representative surveys using 24 h urine collection measurements.[400] Between 2001 and 2011 the mean salt consumption in the UK dropped from 9.5 g/d to 8.1 g/d.[68] This is success, however still far from the national target of 6.0 g/d.[62]

In the UK, salt consumption is higher in more deprived groups.[90, 91] Therefore, interventions which aim to reduce salt consumption should ideally aim to also reduce socioeconomic inequalities in health. Unfortunately, the current UK strategy might potentially increase socioeconomic inequality because awareness campaigns, food labelling, and voluntary reformulation can be more effective among the more health conscious, affluent individuals.[111, 112, 140, 146] Indeed, evidence suggests the socioeconomic gradient in salt consumption might have worsened during the programme.[90, 146] In contrast, modelling studies consistently suggest that more structural interventions can be more effective, cost-effective and equitable than the current UK policy.[176, 401]

Structural salt reduction policies are usually based on legislative initiatives like mandatory reformulation of processed foods or taxation of high salt foods. Such policies have already been adopted successfully in Argentina, South Africa, Portugal, Hungary and elsewhere, emphasising their feasibility. In fact, the actual number of countries currently im-

plementing legislative measures has substantially increased since 2010, indicating a global move towards stricter salt reduction policies.[399]

The aim of this study was to estimate the impact and equity of current UK salt reduction policy on CVD and gastric cancer burden since 2003. I further compared current policy with other feasible policies to estimate possible additional incidence and mortality reductions.

6.2 METHODS

I used the latest IMPACT$_{NCD}$ version as it was described in section 2.2 on page 43 to simulate the effect of current policy and compare it to counterfactual scenarios. I split my analysis into two periods. The first corresponds to years 2003 to 2015, for which I compared the potential benefits of current policy against a null intervention scenario. For the second period, 2016 to 2030, I explored the potential benefits of additional structural salt reduction policies, assuming they might lead to steeper decline in salt intake.

6.2.1 *Period 2003 to 2015 scenarios*

Two scenarios were simulated. The 'no intervention' scenario assumes that no salt related interventions were implemented since 2003. Therefore, the salt exposure remained stable at the estimated level of 2003 for the period up to 2015. The 'current policy' scenario simulated the decline in salt consumption that was observed between 2003 and 2011, and projected it up to 2015 assuming a logarithmic decline.

6.2.2 *Period 2016 to 2030 scenarios*

Here I modelled the potential effect of structural, legislative policies on salt intake, aimed to achieve feasible and ideal targets. First, I modelled a 'current policy' (baseline) scenario where the logarithmic decline observed from 2003 to 2011 was projected up to 2030.

In a 'feasible' target scenario: I assumed that in 2016, policies like mandatory reformulation and/or taxation of high salt foods were implemented and as a result, the mean salt consumption will gradually decline to the national target of 6.0 g/d by 2020 for ages 19 to 64. Due to lack of empirical evidence regarding the magnitude of the impact of such policies on salt, I allowed their target to vary from 5.8 g/d to 7.0 g/d following the PERT distribution (please refer to the footnote in section 2.6.2 on page 63 for a sort description of PERT distribution). The intervention was modelled to be more effective for individuals with higher salt consumption.

In an 'ideal' target scenario: I assumed mean salt intake to reach the ideal salt intake 3.8 g/d by 2025 for ages 19 to 64. The ideal salt consumption was modelled to vary from 1.5 g/d to 6.0 g/d following a PERT distribution. Similarly to the previous scenario, the intervention was modelled to be more effective for individuals with higher salt consumption. The selection of the ideal salt exposure level and its uncertainty was based on published ob-

servational and experimental studies summarised in the appendix Text S4 in Mozaffarian *et al.*[59]

6.2.3 *Salt exposure modelling*

The exposure of the synthetic population to salt and its trend was informed by four nationally representative surveys employing 24 h urine collections between 2001 and 2011.[68, 240–242] I used a stochastic process to enhance the information from these surveys with information from spot urine measurements, as it was described in section 2.3.2.4 on page 51. Then, I used quantile regression to project daily salt consumption to 2030. Changes in salt consumption were transformed to SBP changes using the meta-regression equation from a meta-analysis of 103 trials.[59] The ideal level of salt consumption is not clear (see appendix Text S4 in Mozaffarian *et al.*).[59] I allowed the level of ideal salt consumption under which no risk exists to vary between 1.5 g/d and 6.0 g/d with a mode of 3.8 g/d, following a PERT distribution (table A.1 on page 187).

6.2.4 *Relevant model assumptions*

I assumed a mediated effect through SBP on CVD incidence with 5-year mean lag time and a direct effect to gastric cancer incidence with a mean lag time of 8 years. Furthermore, I assumed that CVD and gastric cancer case fatality is improving by 5 % and 2 % annually, respectively, but the rate of improvement diminishes by 1 % (relative) every year. Finally, I assumed that there is a constant fatality rate socioeconomic gradient of approximately 5 % by QIMD level (halved for ages over 70) forcing the more deprived to experience worse disease outcomes. These assumptions are based on empirical evidence.[21, 92, 93, 109, 261]

Specifically for this study, the no intervention scenario was modelled by stopping the time in 2003 for the quantile regression equation that predicts salt consumption. The impact on SBP salt reduction was estimated by rerunning the same equation for the appropriate year and calculating the difference for each synthetic individual using the formula from Mozaffarian *et al.*[59] The feasible and ideal scenarios were modelled by allowing the current policy scenario to progress up to Step 4 (figure 2.1 on page 45). Then, the mean salt consumption in the population aged 20 to 64 was calculated. From the year the intervention was applied (2015) if the mean was higher than the target, salt consumption of every synthetic individual was multiplied by the target divided by the mean of the synthetic population. Therefore, I applied a proportional reduction to all synthetic individuals and those with higher salt consumption had the higher reduction, in order for the synthetic population mean for ages 20 to 64 to reach the target. The impact of salt reduction on SBP was calculated as in the no intervention scenario. Figure 6.1 on the next page shows the density plots of salt consumption for the scenarios of this study, in one iteration of the simulation.
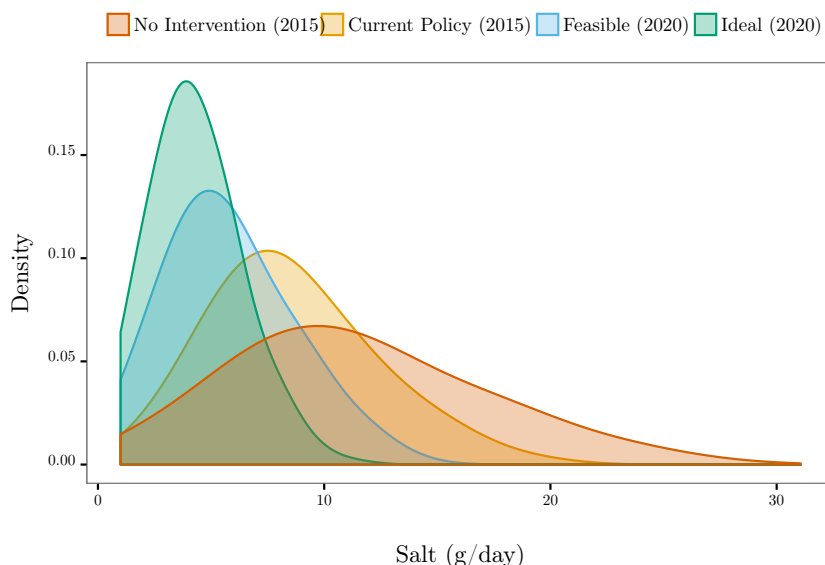
Figure 6.1: Density plot of salt consumption distribution for each scenario of this study in a simulated year. The distributions are truncated on the left because the algorithm does not allow salt consumption <1.0 g/d.

6.2.5   *Model outputs*

IMPACT$_{NCD}$ estimated the cumulative cases prevented or postponed and deaths prevented or postponed from CVD and gastric cancer for the relevant period and for ages 30 to 84. The results were stratified by QIMD. Because of the assumed lag times, any changes in salt exposure in the 2003 to 2015 period are reflected on CVD incidence and mortality in years 2008 to 2020 and gastric cancer incidence and mortality, in years 2011 to 2023. Similarly, for the period 2016 to 2030 these changes are reflected in CVD burden in 2021 to 2035 and in gastric cancer burden in 2024 to 2038.

I summarised the output distributions by reporting medians and IQR in the form of first and third quartiles. I also report the probability (Ps) that a policy scenario aspect is superior to the counterfactual one. For example, '100 cases prevented or postponed (Ps = 80 %) in scenario 'A' is interpreted as 'in 80 % of Monte Carlo iterations at least one case has been prevented or postponed in scenario 'A' comparing to the counterfactual scenario'. Consequently, in the remaining 20 % of iterations, cases in scenario 'A' were more than in the counterfactual scenario. This does not mean that scenario 'A' was harmful, but that its effect in those particular settings was not large enough to exceed the 'noise level' from other sources of uncertainty in the model. To assess the equity of the modelled policies, I used two regression based metrics inspired by the slope index of inequality: the absolute equity slope index; and the relative equity slope index (both described in section 2.7.2 on page 64).

6.3 RESULTS

I have presented my results separately for the two distinct periods.

6.3.1 *Evaluation of current policy (2003 to 2015)*

Under the current policy scenario, median salt consumption was reduced from 8.9 g/d (IQR: 8.7 g/d to 9.2 g/d) in 2003 to 7.1 g/d (IQR: 6.9 g/d to 7.2 g/d) in 2015.[41] Socioeconomic inequalities in salt consumption remained and might even have increased as a result of the current policy.

Under the no intervention scenario IMPACT$_{NCD}$ estimated approximately 1.3 (IQR: 1.2 to 1.4) million new cases of CVD and 700 000 (IQR: 680 000 to 720 000) deaths from CVD. Likewise, the model estimated approximately 68 000 (IQR: 61 000 to 74 000) new gastric cancer cases and 41 000 (IQR: 37 000 to 44 000) deaths.

Compared with the no intervention scenario, the salt reduction strategy resulted in about 52 000 (IQR: 34 000 to 76 000; Ps = 99 %) fewer new CVD cases, and 10 000 (IQR: 3000 to 17 000; Ps = 86 %) fewer CVD deaths. In addition, the current policy prevented around 5000 (IQR: 2000 to 7000; Ps = 92 %) new cases of gastric cancer resulting in 2000 (IQR: 0 to 4000; Ps = 78 %) fewer gastric cancer deaths.

When equity was considered, I estimated that the current policy has a rather neutral effect on tackling socioeconomic inequalities in CVD. The effect on gastric cancer equity was more complex. Current policy apparently prevented or postponed fewer gastric cancer cases in more deprived areas. However, gastric cancer incidence increases with age and more affluent individuals tend to live longer. After directly standardising age and sex, the effect essentially disappeared for absolute inequality but remained for relative inequality (table 6.1 on the next page).

6.3.2 *Future options (2016 to 2030)*

Under the current policy scenario, IMPACT$_{NCD}$ projected that median salt consumption would reduce further from 7.0 g/d (IQR: 6.8 g/d to 7.7 g/d) in 2016 to 6.2 g/d (IQR: 5.9 g/d to 6.2 g/d) in 2030. The addition of structural policies might reach the national target of 6 g/d by 2020. The less feasible ideal policy scenario was estimated to reach 3.6 g/d (IQR: 3.0 g/d to 4.1 g/d) by 2030. Inequality in salt consumption persisted under the current policy projections and decreased moderately with the addition of structural policies.

Under the current policy scenario, I calculated approximately 1.4 million new cases of CVD (IQR: 1.3 to 1.4 million) and 530 000 deaths (IQR: 510 000 to 560 000). Similarly, for gastric cancer I estimated some 80 000 new cases (IQR: 65 000 to 93 000) and 42 000 deaths (IQR: 35 000 to 49 000). Approximately 20 000 more cases of CVD and gastric cancer can be

---

41 The more observant readers may have noticed that these estimates are lower than the estimates from the sodium surveys that I reported in the introduction of this chapter. This is because the survey estimates are for ages 20 to 64, while IMPACT$_{NCD}$ estimates are for ages 30 to 84, and salt consumption decreases with age.

Table 6.1: Effectiveness of current policy compared with the no intervention scenario by fifth of deprivation. The slope for absolute and relative reduction represents the absolute and relative equity slope index, respectively. Brackets contain interquartile ranges for the estimated cases prevented or postponed and 95 % confidence interval (CI) for the slopes.

Deprivation fifth*	Cases prevented or postponed [absolute reduction in thousands]		Cases prevented or postponed [relative percentage reduction]	
	Cardiovascular disease	Gastric cancer	Cardiovascular disease [%]	Gastric cancer [%]
1 (least deprived)	9.7 (4.6 to 16.2)	1.0 (−0.1 to 2.1)	4.1 (1.9 to 6.5)	7.3 (−0.9 to 15.3)
2	11.7 (5.5 to 18.8)	1.1 (0.0 to 2.3)	4.4 (2.3 to 6.8)	7.8 (0.0 to 16.1)
3	11.3 (5.3 to 17.8)	1.0 (−0.2 to 2.0)	4.3 (2.2 to 6.4)	6.9 (−1.3 to 14.7)
4	10.8 (5.0 to 17.5)	0.8 (−0.1 to 1.9)	4.3 (2.1 to 6.7)	6.5 (−1.0 to 15.6)
5 (most deprived)	9.2 (3.8 to 15.5)	0.9 (−0.2 to 2.0)	3.9 (1.6 to 6.0)	7.2 (−2.1 to 15.6)
Slope (crude)	−0.7 (−1.6 to 0.2)	−0.4 (−0.6 to −0.2)	−2.9 (−6.1 to 0.4)	−1.6 (−2.8 to −0.3)
Slope (directly age- and sex-standardised)	4.7 (3.8 to 5.7)	0.2 (0.0 to 0.3)	−0.1 (−0.5 to 0.2)	−1.5 (−2.7 to −0.2)

*According to Index of Multiple Deprivation.
Results are rounded to the nearest tenth.

Table 6.2: Additional cases and deaths that can be potentially prevented or postponed from the addition of structural policies to current policy, and under the ideal scenario compared to the current policy projections for 2015 to 2030. Brackets contain the respective interquartile ranges and the probability of superiority (Ps).

Scenario	Cardiovascular disease		Gastric cancer	
	Cases prevented or postponed in thousands	Deaths prevented or postponed in thousands	Cases prevented or postponed in thousands	Deaths prevented or postponed in thousands
Feasible	18.7 (8.0 to 29.5; Ps = 90 %)	3.6 (-0.4 to 8.1; Ps = 72 %)	1.2 (-0.2 to 3.0; Ps = 72 %)	0.7 (-0.9 to 2.3; Ps = 63 %)
Ideal	73.2 (53.9 to 94.3; Ps = 100 %)	11.0 (6.5 to 16.1; Ps = 95 %)	6.3 (3.4 to 9.6; Ps = 94 %)	3.1 (1.1 to 5.1; Ps = 86 %)

Results are rounded to the nearest tenth.

prevented or postponed from the implementation of structural policies. Table 6.2 presents $\text{IMPACT}_{\text{NCD}}$ estimates for the two counterfactual scenarios in this period.

The addition of structural policies was more effective among the most deprived groups especially for CVD and might potentially decrease absolute socioeconomic inequality (table 6.3 on the next page). As anticipated, the ideal scenario had the largest impact on burden and inequality (table 6.4 on page 143).

Table 6.3: Additional effectiveness of structural policies compared to the current policy scenario by fifth of deprivation. The slope for absolute and relative reduction represents the absolute and relative equity slope index, respectively. Brackets contain interquartile ranges for the estimated cases prevented or postponed and 95 % confidence interval (CI) for the slopes.

Deprivation fifth*	Cases prevented or postponed [absolute reduction in thousands]		Cases prevented or postponed [relative percentage reduction]	
	Cardiovascular disease	Gastric cancer	Cardiovascular disease [%]	Gastric cancer [%]
1 (least deprived)	2.7 (–1.0 to 6.4)	0.3 (–0.7 to 1.1)	1.6 (–0.5 to 3.6)	2.6 (–6.2 to 10.3)
2	2.4 (–1.2 to 6.6)	0.2 (–0.7 to 1.2)	1.3 (–0.7 to 3.6)	2.4 (–6.6 to 10.4)
3	2.8 (–1.0 to 6.8)	0.2 (–0.7 to 1.2)	1.5 (–0.7 to 3.6)	2.4 (–7.0 to 10.2)
4	2.8 (–1.3 to 7.0)	0.2 (–0.7 to 1.0)	1.6 (–0.7 to 3.9)	2.2 (–7.5 to 11.2)
5 (most deprived)	3.3 (–0.9 to 7.3)	0.3 (–0.7 to 1.2)	1.8 (–0.6 to 4.0)	2.7 (–7.7 to 11.6)
Slope	0.6 (0.0 to 1.1)	0.0 (–0.1 to 0.2)	0.2 (–0.1 to 0.5)	0.3 (–1.1 to 1.6)
Slope (directly age- and sex-standardised)	1.7 (1.1 to 2.3)	0.1 (0.0 to 0.2)	0.1 (–0.2 to 0.4)	–0.2 (–1.6 to 1.1)

*According to Index of Multiple Deprivation.
Results are rounded to the nearest tenth.

Table 6.4: Additional effectiveness of the ideal scenario compared to the current policy scenario by fifth of deprivation. The slope for absolute and relative reduction represents the absolute and relative equity slope index, respectively. Brackets contain interquartile ranges for the estimated cases prevented or postponed and 95% confidence interval (CI) for the slopes.

Deprivation fifth*	Cases prevented or postponed [absolute reduction in thousands]		Cases prevented or postponed [relative percentage reduction]	
	Cardiovascular disease	Gastric cancer	Cardiovascular disease [%]	Gastric cancer [%]
1 (least deprived)	7.7 (3.3 to 12.6)	0.8 (−0.3 to 1.7)	4.2 (2.0 to 6.5)	6.7 (−2.7 to 15.2)
2	8.2 (3.6 to 12.6)	0.7 (−0.2 to 1.7)	4.1 (1.9 to 6.2)	5.6 (−1.7 to 14.4)
3	8.9 (4.0 to 14.4)	1.0 (−0.1 to 2.0)	4.4 (2.1 to 6.9)	8.5 (−0.9 to 17.4)
4	8.6 (3.5 to 13.3)	0.7 (−0.2 to 1.6)	4.4 (1.9 to 6.7)	6.8 (−2.0 to 15.8)
5 (most deprived)	9.7 (4.7 to 14.8)	1.0 (0.1 to 1.9)	4.9 (2.5 to 7.1)	9.3 (1.0 to 18.4)
Slope	2.1 (1.4 to 2.8)	0.3 (0.1 to 0.4)	0.8 (0.5 to 1.2)	3.4 (2.0 to 4.7)
Slope (directly age- and sex-standardised)	5.7 (5.0 to 6.3)	0.6 (0.4 to 0.7)	0.7 (0.3 to 1.0)	2.9 (1.5 to 4.3)

*According to Index of Multiple Deprivation.
Results are rounded to the nearest tenth.

6.4   DISCUSSION

This is the first study to quantify the impact of UK salt reduction policies on CVD and gastric cancer by socioeconomic group. I estimated that the current UK salt strategy has potentially prevented or postponed some 57 000 new cases and 12 000 deaths from CVD and gastric cancer in England between the years of 2003 and 2015. The addition of structural policies and achievement of the national target by 2020 could potentially prevent or postpone a further approximately 20 000 new cases and 4000 deaths, while the ideal combination of salt reduction policies might potentially prevent or postpone some 80 000 new cases and 14 000 deaths from CVD and gastric cancer.

When equity is considered, the impact of the implemented strategy is more complex. My results agree with previous studies that the socioeconomic gradient in salt consumption would not be reduced by these strategies.[90, 402] IMPACT$_{NCD}$ estimated that current policies might have a rather neutral impact of CVD socioeconomic inequalities (absolute and relative) and worsen gastric cancer inequalities reflecting an older age distribution in more affluent groups. However, the addition of structural policies may reduce absolute socioeconomic inequality in CVD incidence and neutralise the negative impact of current policies on gastric cancer inequalities.

Simpler modelling studies have previously examined the impact of a theoretical decrease in UK salt consumption. A 3 g/d reduction in salt consumption might prevent about 32 000 CVD cases and 4500 CVD deaths in England and Wales in a 10 year period according to Barton *et al.*, or 200 000 CVD fewer events and 90 000 CVD fewer deaths according to Dodhia *et al.* or almost 100 000 less CVD deaths in 20 years according to Hendriksen *et al.*[131, 403, 404] My results appear to echo the more conservative estimates by Barton *et al.* In addition, the Gillespie *et al.* model informed by experts' opinion estimated that mandatory salt reformulation might reduce socioeconomic inequalities in CHD.[176] I reached reassuringly similar conclusions using a very different methodology.

Going further than previous studies, I modelled structural interventions targeting processed food as being more effective for those individuals with the highest salt intakes. In the UK, about 70 % of dietary salt comes from processed food and is reasonable to assume that those with higher salt consumption have also higher exposure to processed food.[62] Therefore, structural policies targeting processed foods would be more effective for those with higher salt intake.

6.4.1   *Public health implications*

This study confirms and quantifies the positive impact of the currently implemented UK salt reduction policies on CVD and gastric cancer disease burdens. However, I also highlight two culprits of current policy. First, the national target of 6 g/d is unlikely to be reached in the next 15 years assuming the decline continues to be logarithmic. Second, the current policy will probably not reduce socioeconomic inequalities in CVD incidence and might even modestly increase inequalities in gastric cancer. However, structural policies, like mandatory reformulation of processed foods, could potentially accelerate the decline

in salt consumption and also reduce absolute inequality in CVD. The existing salt reduction recommendations for the food industry could achieve the national target.[400] In fact, if the reduction is gradual and consistent, it is unlikely to be noticeable by consumers.[365] In order to realise this however, the food industry must comply with them, which is not happening at present.[405] Failing to do so, will most affect the poorest in society. In addition, the overall impact of this failure is likely to be greater, for example through kidney disease, which I have not considered in this study.

Compared to the results from the previous chapter (chapter 5 on page 121), it is likely that the applied salt reduction policies over the last decade have prevented substantially more CVD cases and deaths than the Health Checks programme will do in the next two decades. This provides further evidence that population-wide interventions can be far more effective than high-risk approaches, even if they do not have strong structural elements. Moreover, my results support the theoretical expectation that the increase of structural elements of a policy may increase both its effectiveness and its equity (section 1.5 on page 30). However, as highlighted in this chapter, in practice the increase in effectiveness and equity depends also on specific characteristics of the population; i. e. the age distributions in the socioeconomic groups. In this particular case because of the different lag times for CVD and gastric cancer prevention and different disease epidemiology, the prevention of CVD deaths in younger ages lead to a small increase in the number of individuals who are exposed to the risk of dying from gastric cancer in older ages. Unlike traditional research methods, IMPACT$_{NCD}$ implements a competing risks framework to explicitly model this phenomenon.

6.4.2   *The salt controversy*

Dietary sodium is necessary to human physiology; therefore, a sodium free diet is infeasible. The exact range of dietary sodium consumption, hence salt consumption, that poses no risk due to inadequacy or excessiveness is not clear. Nevertheless, major institutions like WHO and UK authorities have converged, based on existing evidence, that the recommended salt consumption of 6 g/d is safe. Lately, some researchers argued that salt consumption less than 7.5 g/d may actually increase mortality and a strong polarisation of scientific opinions have been observed.[60, 406] The truth is that these researchers base their argument mainly on studies that used spot and not 24 h urine measurement and then used a mathematical formula to extrapolate daily sodium consumption.[63, 64] This is important because it has been shown that this approach underestimates high sodium intake and overestimates low sodium intake.[215] Therefore, the apparent U shape association between sodium intake and CVD mortality in studies that used spot urine measurements can be explained by measurement bias.[65] However, the U shape association has been also observed in a study that used 24 h urine measurements.[407]

6.4.3   *Strengths and limitations*

This study uses a technically advanced microsimulation model that synthesises information from the best available sources of information on population exposures to salt, and other NCD related risk factors, to generate a 'close to reality' synthetic population. Many assumptions must be made with such models. Yet, in spite of the potential frailty of such assumptions this model validated well against observed CVD and gastric cancer incidence and mortality in real populations, even when multiply stratified. This validation is particularly important because for the years after 2006 the incidence and mortality in the synthetic population were firstly recreated from epidemiological principles and not through an optimisation process. Moreover, to ensure transparency, I have made IMPACT$_{NCD}$ source code open under GNU GPLv3 license.

This study has many limitations, two of which are noteworthy. First, for the evaluation of current policy, I assumed that the decline in salt consumption observed since 2003 was fully attributable to the implemented policy. This was perhaps slightly simplistic, and my estimates may therefore be high. However, this overestimation of the baseline would therefore reduce the apparent gains from additional structural policies, making my conclusions relatively conservative. Second, I could not find a sufficiently large dataset with individual level 24 h urine sodium measurements and other NCD related risk factor information. The stochastic process I developed to overcome this and synthesise information from multiple sources increased overall uncertainty of the model. Nevertheless, this uncertainty has been quantified and transparently reported using uncertainty intervals.

6.5   CONCLUSIONS

Current salt reduction policies are generally effective in reducing the cardiovascular and gastric cancer disease burden but fail to do so equitably. Additional structural policies could achieve further, more equitable health benefits.

In the next chapter, I will explore tobacco control policies that, like salt policies, can prevent multiple NCDs. Moreover, I will enrich the spectrum of prevention typologies that IMPACT$_{NCD}$ can model by simulating fully structural, tobacco endgame policies.

TOBACCO: THE ENDGAME?

7.1 INTRODUCTION

In the previous chapters, I focused on existing primary prevention policies and their incremental improvements. In this chapter, I will explore a more ambitious option: a tobacco sales ban. Although the effectiveness of a potential tobacco sales ban is unquestionable, its feasibility in the current political context is contested. I will discuss some aspects of feasibility of a tobacco sales ban; however, my intention is to concentrate more on the dynamics of the phenomenon and how a sharp decline in smoking prevalence could reflect on the future burden of CVD and lung cancer over time.

Globally, approximately 5.8 million deaths could be attributed to smoking in 2013, an increase of 1.2 million since 1990. Environmental tobacco smoking accounted for additionally more than 330 000 deaths in 2013. Tobacco is the leading cause of DALYs[42] in most high-income countries and among the leading risks overall.[33]

In England, tobacco is the leading cause of DALYs among women and the second leading cause, surpassed only by unhealthy diet, among men. Almost 11 % of DALYs were lost due to smoking in 2013.[20] During the same year, smoking caused an estimated 80 000 deaths in England among adults aged 35 and over. This amounts to 17 % of all deaths for these ages, and has been unchanged since 2005. Over 450 000 hospital admissions were attributable to smoking, representing 4 % of all adult admissions.[408] These numbers may underestimate the true burden of smoking, as a recent study has expanded the list of diseases linked to smoking.[41]

Despite the undeniable risk of smoking to health, smoking remains common in England, with 19 % of adults aged 16 and over reported as smokers in 2013.[409] This prevalence fell slightly from 21 % in 2007.[43] Furthermore, large differences in smoking prevalence persist across socioeconomic groups; over 30 % of people with routine and manual jobs smoke, compared to less than 15 % of those in managerial and professional occupations.[43, 410, 411] Smoking explains more than one quarter of the socioeconomic gradient in total mortality in Great Britain.[20, 412]

The UK has strong tobacco control policies compared to many European peers, achieving the highest score on the Tobacco Control Scale (74 out of 100) among 34 European countries.[413] The Tobacco Control Scale is an expert developed instrument for assessing the strength of tobacco control policies with data compiled via a survey of national representatives to the European Network for Smoking and Tobacco Prevention, supplemented with data from other data sources (described in more detail in [413]). WHO uses a different ranking system (MPOWER), where the UK rank is reassuringly similar.[414]

---

42 The disability adjusted life year is a combined metric of mortality and morbidity. It measures overall disease burden, expressed as the number of years lost due to ill health, disability or early death.

A modelling study using the IMPACT policy model estimated that if the UK implements tobacco control policies to maximise the Tobacco Control Scale, this could reduce smoking prevalence by about 3 % absolute (15 % relative), prevent more than 3000 premature CHD deaths, and reduce socioeconomic inequalities in health by 2025.[415] Levy *et al.* used the SimSmoke model to model the UK's full compliance with the MPOWER framework, to produce comparable results.[173] Both studies provide evidence that increments of existing tobacco control policies cannot eliminate smoking in the near future.

Many experts have also realised that the policy debate needs to move from 'tobacco control' to 'tobacco free populations', especially for countries with well implemented tobacco control policies. A paradigm shift that, in the jargon of the tobacco control community, is known as the 'endgame'. The endgame may require novel and radical approaches to tackle the tobacco epidemic.[416] Indeed, a full supplement issue of the Tobacco Control journal (May 2013, Volume 22, suppl 1) was dedicated to these innovative proposals that may achieve a drastic reduction in smoking prevalence. More recently, McDaniel *et al.* published a synthesis of the proposed endgame policies so far.[417] They categorise the policies into five large groups:

1. those aimed to the product itself (i. e. regulate nicotine levels to make cigarettes non-addictive or less addictive, redesign the cigarette to make it unappealing, electronic cigarettes);

2. those aimed to the user (i. e. smoker's license, prescription to purchase tobacco, restrict sales by year born);

3. those aimed to the market/supply side (i. e. licensing, outlet restrictions, display bans and price controls, ban combustibles, advantage cleaner nicotine products over combustibles, 'sinking lid', price caps);

4. institutional structure focused (i. e. tobacco control agency, regulated market model, state takeover of tobacco companies, performance based regulation);

5. integrated endgame strategies (i. e. combination of existing and increment tobacco control policies with one or more of the endgame strategies).

From the policies mentioned above, the most feasible and mature appear to be the 'tobacco free generation' by restricting sales by year born (group 2 in previous list),[418, 419] and the total ban of combustibles (group 3 in the previous list).[420, 421] The former is under discussion in the local government of Tasmania[422] and is backed by the British Medical Association in the UK[423], while the latter has already been implemented in Bhutan and soon to be in Turkmenistan.[424, 425]

The aim of this chapter is to explore the dynamics of two endgame policies on the burden and socioeconomic inequalities of CVD and lung cancer. The endgame policies that were simulated, were the 'total ban of combustibles' and the 'tobacco free generation'.

7.2   METHODS

I used the final version of IMPACT$_{NCD}$ as it was described in the methods section (section 2.2 on page 43). The synthetic population was representative of the 2006 community dwelling English population and the simulation horizon was set to 40 years, up to 2045.

7.2.1   *Scenarios*

Four scenarios were simulated: the 'current policy' scenario; the 'total ban' scenario; the 'tobacco free generation' scenario; and the 'ideal minimum'.

CURRENT POLICY SCENARIO: Similarly to the previous result chapters, this scenario assumes that the recent observed risk factor trends, likely driven by current tobacco policy efforts, will continue in the future.

TOTAL BAN SCENARIO: This scenario simulates a total ban of combustibles by 2016. In other words, the sale of tobacco combustible products (like cigarettes and cigars) is prohibited by law, and the state has the means to enforce the law adequately. I assumed a 50 % reduction in smoking initiation rate, a 50 % reduction in active to ex-smoking ratio, and a 50 % reduction in cigarette consumption compared to the 'current policy' scenario. These estimates were roughly based on data from Bhutan, the only country where a tobacco ban has been implemented so far. In Bhutan, four years after the ban about 10 % of men and 7 % of women were active smokers mostly reflecting an expected rise in tobacco products smuggling.[426]

TOBACCO FREE GENERATION SCENARIO: This scenario models a ban on the sale of tobacco products to anyone born in or after 2000. Similarly to the previous scenario, I assumed a 50 % reduction in smoking initiation rate, and a 50 % reduction in active to ex-smoking ratio for the synthetic individuals born in or after 2000. I assumed no reduction in cigarette consumption to reflect a higher availability of tobacco products compared to the previous scenario. Hence, I assumed those who may have illegal access to tobacco products will not reduce their consumption.

IDEAL MINIMUM SCENARIO: This is a theoretical scenario that assumes the population had never been exposed to smoking. It can be used as a marker of the total burden of CVD and lung cancer attributable to smoking.

7.2.1.1   *Common scenario assumptions*

I assumed that the CVD case fatality rate is improving by about 4.5 % and the lung cancer case fatality rate is improving by 3 % annually. I also assumed that the case fatality rate of improvement gradually declines to avoid a very low case fatality rate in later years of the simulation. Moreover, I assumed that there is a constant fatality rate socioeconomic

gradient of approximately 8 % by QIMD level (halved for ages over 70) forcing the more deprived to experience worse disease outcomes. These assumptions were based on empirical evidence.[21, 92, 93, 107, 108]

7.2.2   *Model outputs*

IMPACT$_{NCD}$ estimated the cases prevented or postponed and deaths prevented or postponed from CVD and lung cancer between years 2016 and 2045 and for ages 30 to 84. The results were stratified by QIMD. I summarised the output distributions by reporting medians and IQRs in the form of first and third quartiles. I also reported the probability (Ps) that a policy scenario aspect is superior to the current practice one. To assess the equity of the modelled policies, I used two regression based metrics inspired by the slope index of inequality: the absolute equity slope index; and the relative equity slope index (both described in section 2.7.2 on page 64).

7.2.3   *Model alignment*

During the validation process of IMPACT$_{NCD}$ (chapter 3 on page 77), it was revealed that IMPACT$_{NCD}$ underestimated the observed upward trend in lung cancer incidence in woman (figure 3.1 on page 80). A closer inspection revealed that the underestimation occurred specifically in women aged 70 to 74. I aligned[43] the model by inflating the number of cigarettes smoked in this particular age group by 10 %. Moreover, the model overestimated lung cancer mortality for older ages in the more deprived groups (figure 3.12 on page 92 and figure 3.13 on page 93). I chose not to align this for simplicity. This choice added negligible bias to the results, because IMPACT$_{NCD}$ is primarily an incidence model, and the effectiveness and equity of interventions are estimated from cases and not deaths.

7.3   RESULTS

I first present the cumulative cases and deaths prevented or postponed for the total simulation period between 2016 and 2045. Then, I present the results by year to capture and analyse the impact of the policies, dynamically.

7.3.1   *Smoking prevalence*

The active smoking prevalence between 2016 and 2045 was predicted to reach 14.8 % (14.4 % to 15.1 %) for men and 8.8 % (8.6 % to 9.1 %) for women under the current policy. Under the total ban scenario the active smoking prevalence was predicted to fall to 7.2 % (7.0 % to 7.4 %) for men and 4.3 % (4.1 % to 4.5 %) for women. Finally, under the tobacco free generation scenario prevalence was estimated to 12.4 % (12.1 % to 12.6 %) for men and 7.9 % (7.7 % to 8.1 %) for women.

---

43 Alignment is the term that is widely used to refer to the calibration process of microsimulations.

7.3.2   *Disease burden*

IMPACT$_{NCD}$ estimated approximately 2.6 (2.3 to 2.7) million CVD cases and some 750 000 (740 000 to 760 000) lung cancer cases between 2016 and 2045 under the current policy scenario. These would result in approximately 1.0 (0.9 to 1.1) million CVD deaths and some 610 000 (600 000 to 620 000) lung cancer deaths over the same period.

Under the total ban scenario about 90 000 (70 000 to 120 000, Ps = 99.9 %) CVD cases and 79 000 (55 000 to 100 000, Ps = 99.5 %) lung cancer cases were prevented or postponed. This resulted in 14 000 (3000 to 25 000, Ps = 80.5 %) fewer CVD deaths and 54 000 (38 000 to 73 000, Ps = 99.0 %) fewer lung cancer deaths.

Under the tobacco free generation scenario about 3500 (−4200 to 11 000, Ps = 62.8 %) CVD cases were prevented or postponed, resulting in about 190 (−2900 to 3400, Ps = 51.8 %) fewer deaths. Furthermore, IMPACT$_{NCD}$ estimated approximately 230 (−3100 to 3600, Ps = 51.8 %) less lung cancer cases and 220 (−2700 to 3200, Ps = 52.4 %) deaths compared to the baseline scenario.

Finally, under the ideal minimum scenario approximately 350 000 (310 000 to 390 000, Ps = 100 %) CVD cases were prevented or postponed, resulting in about 58 000 (45 000 to 71 000, Ps = 100 %) fewer deaths. Additionally, IMPACT$_{NCD}$ estimated approximately 660 000 (650 000 to 680 000, Ps = 100 %) less lung cancer cases and 520 000 (510 000 to 540 000, Ps = 100 %) deaths compared to the baseline scenario.

7.3.3   *Policies equity*

All scenarios were estimated to be substantially equitable and can potentially have a major impact on reducing absolute and relative socioeconomic inequalities in CVD and lung cancer (table 7.1 on the next page and table 7.2 on page 153). The ideal minimum scenario specifically, is a marker of the smoking attributable socioeconomic inequalities in health; the inequitable tobacco control policies that have been implemented in recent years have also contributed to these inequalities (please refer to section 1.5.4 on page 33).

7.3.4   *Policy dynamics*

When the effect of each scenario is analysed by year (figure 7.1 on page 155), it is apparent that the health impact of the tobacco free generation policy would be small. This can be easily explained by the age distribution of the tobacco free generation that would be maximum 45 years in 2045; hence, the potential of prevention in these age groups is limited because the burden of CVD and lung cancer is small anyway. That said, I expect the potential benefits of this strategy to materialise within the next four decades since the simulation horizon.

On the other hand, the total sales ban may yield more timely results especially for CVD. CVD cases prevented or postponed will gradually decrease after the initial increase from the policy implementation reflecting the favourable trends of other CVD related risk factors. On the contrary, lung cancer cases prevented or postponed are predicted

Table 7.1: Cases prevented or postponed according to fifth of deprivation by 2045, along with the absolute equity slope index for each scenario. Brackets contain interquartile ranges for cases prevented or postponed and confidence intervals for the equity slope index.

Deprivation fifth*	Total ban scenario		Tobacco free generation scenario		Ideal minimum scenario	
	CVD [Thousands]	Lung cancer [Thousands]	CVD [Thousands]	Lung cancer [Thousands]	CVD [Thousands]	Lung cancer [Thousands]
1 (least deprived)	4.2 (−3.7 to 13.2)	0.6 (−3.8 to 4.8)	0.1 (−3.3 to 3.2)	0.2 (−1.2 to 1.3)	26.8 (17.0 to 36.0)	71.3 (67.9 to 74.6)
2	12.9 (3.3 to 22.0)	9.2 (3.8 to 15.2)	0.4 (−3.4 to 4.3)	0.0 (−1.5 to 1.4)	52.1 (40.1 to 62.8)	108.1 (103.9 to 112.3)
3	14.3 (3.7 to 24.1)	11.8 (5.3 to 17.3)	0.5 (−3.4 to 4.5)	−0.1 (−1.7 to 1.7)	58.8 (49.1 to 70.7)	129.0 (124.8 to 133.3)
4	21.2 (11.9 to 31.2)	16.1 (9.0 to 23.5)	0.9 (−3.6 to 4.9)	0.1 (−1.7 to 1.7)	83.5 (71.4 to 95.1)	149.9 (145.1 to 154.5)
5 (most deprived)	41.0 (30.3 to 53.4)	41.8 (31.7 to 51.8)	1.7 (−3.5 to 6.6)	0.0 (−2.0 to 2.2)	129.1 (113.2 to 146.3)	202.3 (195.2 to 209.6)
Absolute equity slope index	41.7 (40.1 to 43.2)	44.8 (43.6 to 45.9)	1.8 (1.2 to 2.4)	0.1 (−0.1 to 0.4)	119.4 (117.4 to 121.4)	152.4 (151.3 to 153.4)

*According to Index of Multiple Deprivation.
CVD denotes cardiovascular disease. Results rounded to one decimal place.

Table 7.2: Relative percentage reduction in cases prevented or postponed according to fifth of deprivation by 2045, along with relative equity slope index for each scenario. Brackets contain interquartile ranges for relative reductions and confidence intervals for the equity slope index.

Deprivation fifth*	Total ban scenario		Tobacco free generation scenario		Ideal minimum scenario	
	CVD [%]	Lung cancer [%]	CVD [%]	Lung cancer [%]	CVD [%]	Lung cancer [%]
1 (least deprived)	0.9 (−0.9 to 3.0)	0.7 (−4.4 to 5.4)	0.3 (−6.0 to 5.8)	1.4 (−15.8 to 15.7)	6.0 (3.8 to 7.9)	80.1 (78.1 to 82.0)
2	2.6 (0.6 to 4.5)	7.3 (2.8 to 11.6)	0.5 (−4.5 to 5.5)	0.0 (−16.0 to 13.3)	10.5 (8.6 to 12.6)	85.1 (83.7 to 86.6)
3	2.9 (0.8 to 4.8)	8.0 (3.7 to 11.6)	0.7 (−3.8 to 4.9)	−0.5 (−13.6 to 11.7)	11.9 (9.9 to 13.9)	87.5 (86.3 to 88.5)
4	4.3 (2.4 to 6.1)	9.6 (5.6 to 13.9)	0.8 (−3.7 to 4.6)	0.4 (−9.8 to 11.1)	16.8 (14.8 to 19.0)	89.3 (88.4 to 90.4)
5 (most deprived)	7.4 (5.6 to 9.1)	19.0 (14.8 to 22.9)	1.3 (−2.6 to 4.8)	0.1 (−8.9 to 8.5)	23.1 (21.1 to 25.1)	91.8 (90.9 to 92.6)
Relative equity slope index	7.2 (7.0 to 7.5)	19.7 (19.0 to 20.3)	1.4 (0.6 to 2.1)	3.1 (1.1 to 5.1)	20.5 (20.1 to 20.8)	13.9 (13.7 to 14.1)

*According to Index of Multiple Deprivation.
CVD denotes cardiovascular disease. Results rounded to one decimal place.

to increase gradually during the simulated period reflecting the accumulation of risk from smoking and the longer lag times. The cases prevented or postponed from the ideal minimum scenario were also plotted on the same graph to mark the total cases attributable to smoking.

Figure 7.2 on page 156 and figure 7.3 on page 157 depict the equity impact of the scenarios over time. In general, the effectiveness of the interventions increased with deprivation. Specifically, the total sales ban of tobacco products has a great potential to reduce both absolute and relative socioeconomic inequalities in CVD and lung cancer.
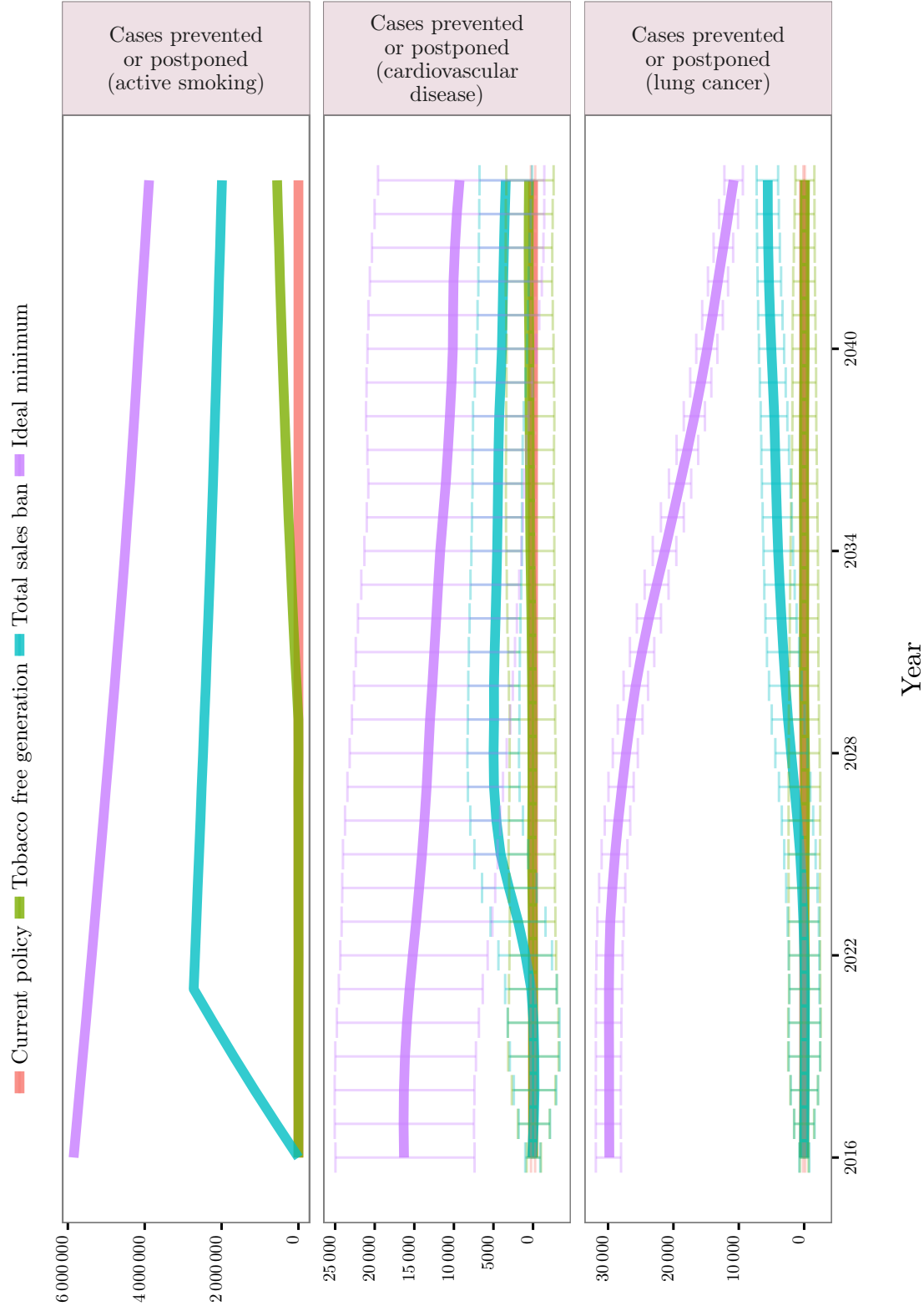
Figure 7.1: Smoking, cardiovascular disease, and lung cancer cases prevented or postponed under the three scenarios, between 2016 and 2045. Error bars depict interquartile ranges.
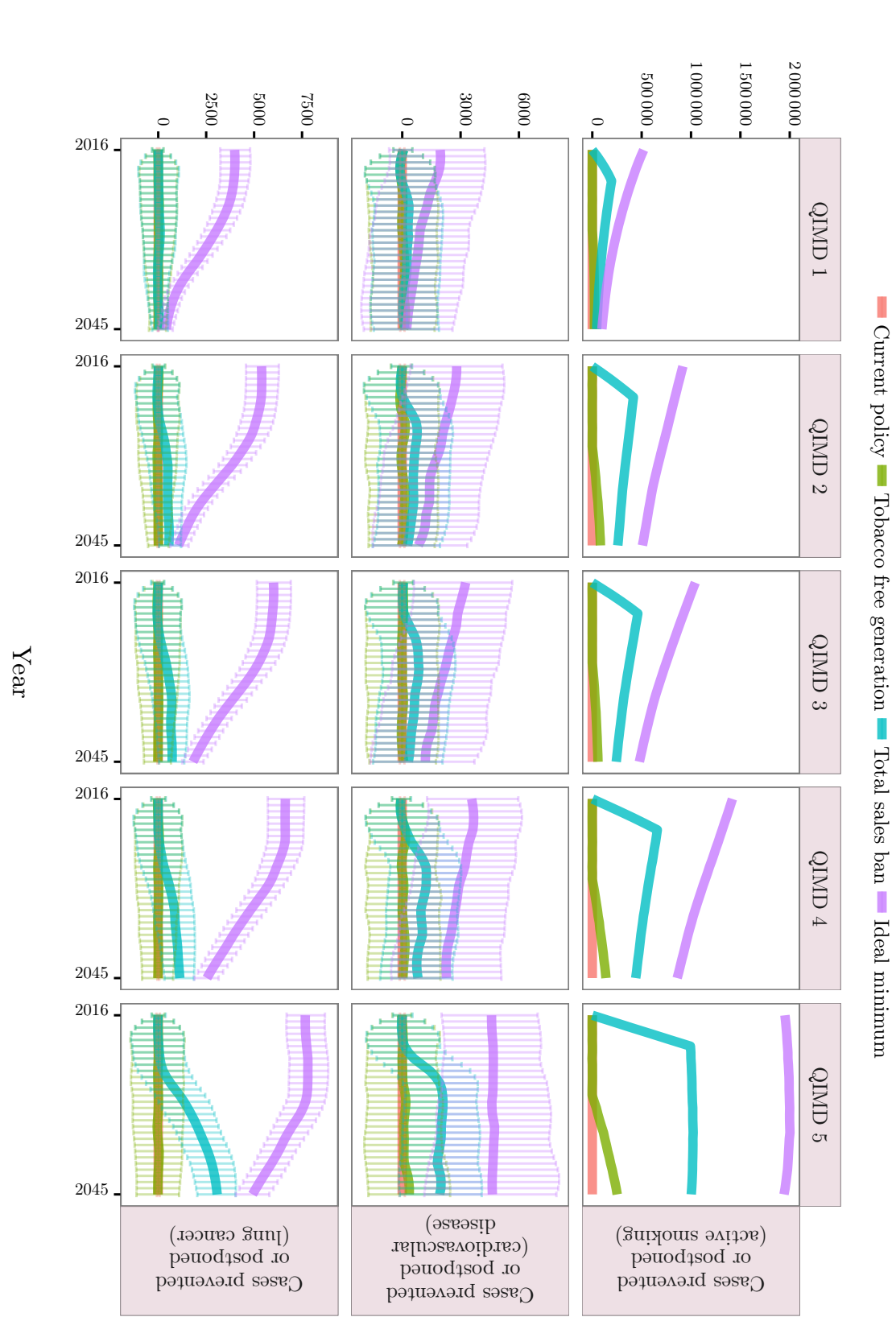
Figure 7.2: Smoking, cardiovascular disease, and lung cancer cases prevented or postponed under the three scenarios, between 2016 and 2045 by fifth of deprivation based on quintile groups of Index of Multiple Deprivation (QIMD) (1 = least deprived, 5 = most deprived). Error bars depict interquartile ranges.
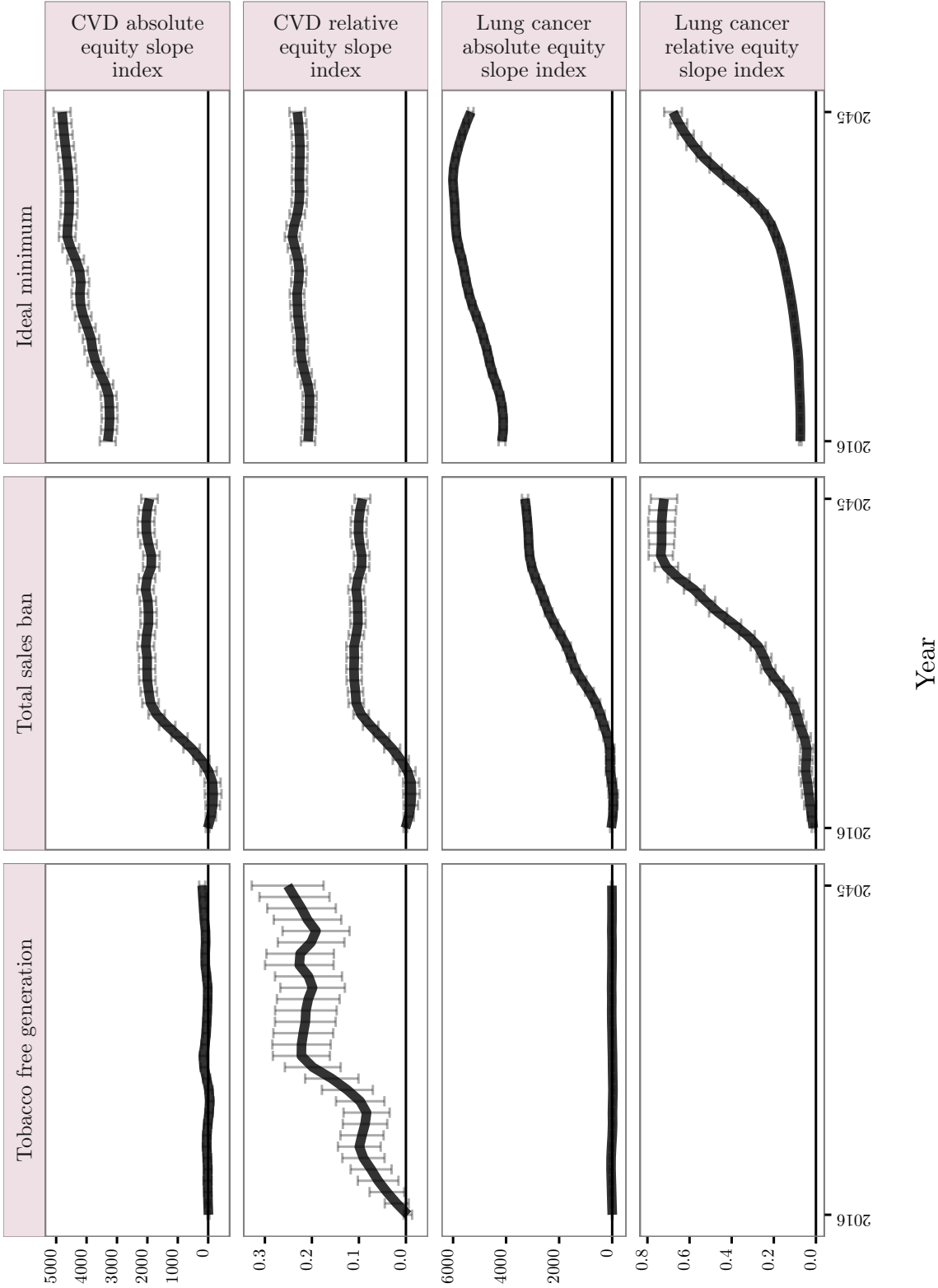
Figure 7.3: Absolute and relative equity slope indices under the three scenarios, between 2016 and 2045. Relative equity slope index for lung cancer under the tobacco free generation scenario was not able to be estimated due to small number of prevented cases each year. Error bars depict interquartile ranges.

IMPACT$_{NCD}$ estimated that the total ban might prevent approximately 170 000 cases of CVD and lung cancer in 30 years, resulting in some 70 000 fewer deaths. The probability of this policy to be more effective than current policy was almost 100 %. The health gains from the policy would first come from CVD prevented cases and about 10 years later the impact on lung cancer would follow, gradually increasing over time. On the other hand, the impact of the tobacco free generation policy was estimated to be substantially smaller; only about 3500 CVD cases and a few hundred deaths could be prevented or postponed from the policy within 30 years and the impact on lung cancer would be negligible. This is because only young individuals, for whom disease burden is small anyway, would be affected by this policy within the simulation horizon. Subsequently, the probability of superiority for this policy was very low. Finally, IMPACT$_{NCD}$ estimated that both policies are equitable, but given the much higher overall effectiveness of a total ban, adoption of a total ban could dramatically reduce socioeconomic inequalities in health.

In March 2016, Cancer Research UK and the UK Health Forum published a report about the potential health and economic benefits of a tobacco free UK society. In their report, they defined a tobacco free society as having smoking prevalence of less than 5 % across the socioeconomic spectrum and they used a dynamic microsimulation to estimate the potential prevented disease cases and costs from achieving this. They estimated that achieving a tobacco free UK may lead to 97 500 fewer disease cases including 36 000 cancer cases, by 2035. These estimates include not only CVD and lung cancer cases, but also cases from other smoking related cancers and chronic obstructive pulmonary disease.[427]

At first glance, these results appear less optimistic than mine, considering that they are for the entire UK and include more diseases than my estimates. Nevertheless, the difference can be explained by the different scenario assumptions in the two models. First, the simulation horizon for IMPACT$_{NCD}$ was 30 years compared to 20 years for the Cancer Research UK and the UK Health Forum report. Second, IMPACT$_{NCD}$ uses HSE data to project future smoking prevalence for the baseline scenario, while the model for the report uses data from the General Lifestyle Survey. This resulted in lower forecasted smoking prevalence for the baseline scenario in the latter; therefore, more conservative counterfactual scenario estimates. Third, I modelled the total ban to rapidly decrease smoking prevalence within five years from implementation. The tobacco free society scenario, in the report, assumed a gradual decrease over the 20 years of the simulation horizon. Finally, structural differences in the two models (i. e. the model of the report ignores lag times and does not model smoking intensity) and the modelled scenarios may have also contributed to the difference of their estimates.

7.4.1 *The scenarios*

From the available options for the tobacco endgame that have been reported in the literature, I chose the two options that have already been implemented or are discussed for implementation. Both policies, the total ban of combustible tobacco products and the

tobacco free generation, would require major legislative, institutional, and bureaucratic changes in order to succeed.[428] Yet, the implementation of a total ban on tobacco sales in Bhutan is a proof of concept that the policy can perhaps be implemented in the UK. It is true that the implementation in Bhutan has been poorly evaluated, led to increasing in tobacco products smuggling, and increased the prosecution of smokers.[424, 426] However, there are lessons to be learned from Bhutan and future implementation may be less problematic, especially in high-income countries with better administrative and governance systems in place. In fact, the recent decision of the UK to leave the European Union may give more flexibility to the UK government to apply such policies and better secure the borders to prevent tobacco smuggling. Proposing a complete and detailed tobacco ban policy is far beyond the scope of this thesis. However, the results of this chapter suggest that the cost in human lives of not including endgame policies in the policy dialogue is substantial. In the future, when the implementation of such policies in England will be better defined and operationalised by experts, modelling exercises may explicitly consider the unintended consequences of endgame policies like the criminalisation of smokers or the promotion of tobacco products smuggling. Finally, I avoided modelling the existence of a lower limit of achievable smoking prevalence, since recent evidence contradict the 'hardening hypothesis'.[44][429, 430]

Critics may argue that bans reduce autonomy, and have proven ineffective in the past. The reduced autonomy argument is present in many public health interventions, and the debate has recently been summarised by Capewell and Lilford.[431] The UK already bans other substances like heroin or 'ecstasy', many of which are less harmful than tobacco and despite the subsequent criminalisation of the users.[432] Therefore, a potential ban of combustible tobacco products is consistent with current policies on other harmful and addictive substances. Going a step further, Proctor proposes that the sales ban of harmful, addictive substances does not restrict but rather enables choice, because it liberates the user from the addiction.[420]

Finally, regarding effectiveness, bans have failed in the past when they were not supported by the majority of the population; with the alcohol 'prohibition era' in the US being the most notable example. In Europe, about 40 % of non-smokers and 20 % of smokers support a total tobacco sales ban.[433, 434] In England, public support is higher by about 5 %.[435] These numbers may seem low at first, but considering there was no public debate or any public awareness campaigns about a tobacco sales ban, they could also represent a promising start.[436]

7.4.2 *Public health implications*

Realistically, public health practitioners and policy makers seek answers in two very pragmatic and interrelated questions. When is the right time to publicly set a target for smoke free societies, and what options exist to reach this target? The first question has been under discussion among public health experts over the last few years. Of course the ques-

---

44 The hardening hypothesis posits that as smoking prevalence declines, the remaining smokers are harder to quit because they are deeply addicted to nicotine.

tion is relevant for countries that have already achieved low smoking prevalence through effective tobacco control policies and in some of these countries (i. e. New Zealand and Finland) a target has been set for a smoke free society. Scotland has also set a target to achieve smoking prevalence of less than 5 % by 2034. In England and the other UK countries public health advocates are trying to involve policy makers in the debate.[427, 436, 437]

My study provides useful evidence that it is highly unlikely that smoking prevalence will be less than 5 % within the next three decades without additional tobacco control efforts. Other modelling studies have shown that even by maximising the effectiveness of existing tobacco control policies it would be unlikely to achieve a 5 % prevalence.[173, 415, 438] Therefore, any discussion about achieving a smoke free society within the next three decades must involve radical tobacco control policies. Furthermore, my study highlights that when considering endgame policies, targeting smoking initiation alone would be inadequate to reduce the burden of smoking in the foreseeable future. Therefore, any realistic endgame plan must involve tobacco control policies targeting current smokers in addition to never-smokers.

In my analysis I included the ideal minimum scenario to quantify and explore the attributable to smoking burden of CVD and lung cancer. The main message from this scenario is that even after a successful dramatic reduction in smoking prevalence the consequences from the cumulative effect of smoking on health would last for decades after. Yet, any delay to substantially reduce smoking prevalence translates into thousands of smoking attributable deaths that mostly burden the more socioeconomically deprived.

Finally, it is worth noting that all the proposed endgame policies are strongly structural or inversely agentic. By inversely agentic I mean that they require smokers to actively engage and mobilise their resources in order to get access to tobacco products. Hence, interestingly the expected attrition of these policies is the actual intervention.

7.4.3   *Strengths and limitations*

Any public health expert would expect that a tobacco sales ban would have a tremendous effect on disease burden and socioeconomic inequalities, while the tobacco free generation policy would need a prolonged period before it yields substantial health benefits to the population. An epidemiologist would also expect the policy effect on CVD burden to be observable earlier than the effect on lung cancer because of the different lag times. However, policy makers are rarely public health experts or epidemiologists. Hence, in this case IMPACT$_{NCD}$ integrates current epidemiological knowledge about tobacco, CVD, and lung cancer and makes explicit assumptions to allow non-experts to understand better the effectiveness, equity, and dynamics of the proposed policies. In addition, this study may be helpful also for experts, because it quantifies the likely effect and equity of the modelled policies.

The generic limitations and assumptions of the model are discussed in other chapters (section 8.5 on page 173 and table 2.1 on page 69). The main limitation of this specific chapter is that it modelled only two of the diseases smoking is associated with, underes-

timating the health impact of the modelled policies. However, CVD and lung cancer are the two smoking related diseases with the highest societal burden, and other diseases can be modelled using the IMPACT$_{NCD}$ framework in the future. The second limitation of this study is that it modelled only cigarette smoking. The prevalence of exclusive non-cigarette smoking (pipes or cigars) is small in the UK, therefore the impact of the modelled policies is slightly underestimated. The exclusion of electronic cigarettes from the simulation though, is perhaps more important and its impact on my estimates is largely unknown. The electronic cigarette is a relatively new product and there is not enough information to allow reasonable prediction of its utilisation over the next 30 years. Nevertheless, the academic debate about the potential role of electronic cigarettes in smoking cessation and smoking or vaping initiation is heated and IMPACT$_{NCD}$ may contribute to the debate in the near future.

7.5 CONCLUSIONS

IMPACT$_{NCD}$ estimated that despite the gentle decrease in smoking prevalence, the consequences of smoking on health and health equity will last for decades. Additional efforts and new radical approaches may be required to achieve a dramatic reduction in smoking prevalence that will target simultaneously smoking initiation and smoking cessation. Discussions about when it is the right moment to set a target for a smoke free England needs to be done in parallel with discussions about how the target may be achieved and what additional tobacco control policies might be required for this.

This chapter concludes the results of my thesis. In the next chapter I will discuss and reflect on the overarching themes that have emerged in my thesis, and the potential role of modelling in informing public health policy.

GENERAL DISCUSSION

8.1 INTRODUCTION

In this chapter I will first summarise my findings in relation to the aims and objectives of my thesis. Then, I will discuss the role of simulation modelling in public health policy and decision-making, and I will highlight some of the key challenges. I will discuss the key assumptions and limitations in IMPACT$_{NCD}$, and I will summarise my future plans. Finally, I will reflect on my research experiences over the last three years.

8.2 KEY FINDINGS WITH REFERENCE TO MY AIMS

The primary aim of my thesis was to construct and validate a simulation model for public health policy that is reusable, transparent, and comprehensive. Then, to use this model to quantify the impact of existing and hypothetical counterfactual primary prevention policies on disease burden and health inequalities (section 1.7 on page 40).

I created and validated IMPACT$_{NCD}$ and then I used it in my four results chapters to explore a wide range of scenarios. I applied the same IMPACT$_{NCD}$ framework to analyse the contribution of statins in the observed cholesterol decline and their equity (chapter 4); the effectiveness and equity of universal and concentrated screening for CVD (chapter 5); the effectiveness and equity of several population-wide approaches (chapters 5, 6, and 7); and the dynamics of two tobacco control policies over time (chapter 7). The policy scenarios I modelled in my thesis cover the full spectrum of primary prevention typologies (please refer to section 1.5 on page 30); from agentic interventions like NHS Health Checks, to population-wide interventions with weak structural components like the current salt reduction strategy, to strong structural policies like mandatory salt reformulation and a tobacco sales ban, to proportionate universalism through combinations of these policies. In addition, I simulated the potential impact of these policies on multiple NCDs simultaneously and with different lag times in a competing risk framework. These arguments support that IMPACT$_{NCD}$ framework is truly reusable and can be used in the future to model more policy scenarios, involving more diseases and different socioeconomic classifications (please refer to section 8.6 on page 176).

IMPACT$_{NCD}$ is also transparent. All data sources are presented and properly referenced. The structure and all the algorithms of the model are described in a non-technical manner and are linked to fundamental epidemiological principles. All the assumptions are explicitly described and when the assumptions are based on empirical evidence such evidence is referenced. Moreover, the modelled scenarios are described in a clear, accessible way, and the justification for the modelling decisions is also provided. Most importantly, the source code of IMPACT$_{NCD}$ is open meaning that anyone can have access to it, use it as

it is, or improve it. The model can be scrutinised not only during the peer review process for publication purposes, but continuously and by anyone that has the relevant technical skills. Additionally, IMPACT$_{NCD}$ code base can be reused from other modellers in part or in full, without any extra permissions as long as the code remains open. This may improve the quality, and reduce necessary resources of other models in this field. To conclude, my approach offers ultimate transparency for technical and non-technical readers; unfortunately, this is rare in public health modelling, despite the well-recognised need for transparency.[161, 186, 287]

Finally, IMPACT$_{NCD}$ is more comprehensive relative to other models. The interplay of socioeconomic inequalities, exposure to risk factors, and disease patterns and trends are in the core of IMPACT$_{NCD}$ by design. Ideas from the social determinants of health framework (section 1.4.1 on page 24) have been infused in the synthetic population of IMPACT$_{NCD}$. The socioeconomic position of synthetic individuals influences their behavioural exposures, and the latter influence their biological exposures. This allowed for the 'social production of disease' framework (section 1.4.2 on page 26) to be incorporated in IMPACT$_{NCD}$, and the processes that generate health inequalities from socioeconomic inequalities to be modelled. Additionally, population trends, risk factor exposure trends, and case fatality trends are also modelled by fifths of deprivation. Hence, IMPACT$_{NCD}$ allows for a more comprehensive and dynamic exploration of the effectiveness and equity of the modelled policies. Despite the limitations of the current implementation of IMPACT$_{NCD}$ that may limit the comprehensiveness of the model (please refer to section 8.5 on page 173), its expandable modelling framework may relax some of these limitations in the future (please refer to section 8.6 on page 176).

An interesting side effect of the increased comprehensiveness of the model, is that the ratio of cases to deaths prevented or postponed varies extensively. For example, it fluctuates from around 9:1 for the combination of population-wide policies with targeted CVD screening in chapter 5 (table 5.1 on page 126), to around 5:1 for the feasible legislative salt reduction policies in chapter 6 (table 6.2 on page 141). This is because different policies prevent disease from individuals with different characteristics. The case fatality rate in the model depends on the type of CVD disease (CHD or stroke), age, sex, socioeconomic status, and calendar year. Consequently, the ratio of cases to deaths prevented or postponed is influenced by the same factors.

8.2.1 *Could I have achieved the aims using different modelling methodology?*

Since I created the list of models in chapter 1 that have been used in multiple studies and have likely informed policy (section 1.6.1 on page 35), at least two more models have emerged to fit these criteria. The first is the PRIME model (an evolution of the DIETRON model).[172, 439, 440] PRIME is a comparative risk assessment model that allows the user to input counterfactual risk exposure distributions that could have occurred by the modelled intervention and quantify the impact of the intervention on disease-specific mortality. It includes multiple risk factors and multiple diseases and it appears it could be easily expanded to model different socioeconomic groups. Therefore, it is reasonably reusable.

However, it is not open source; therefore, other researchers cannot use parts of this model in their modelling approaches and they cannot improve the existing model without prior permission from the authors. The inputs, logic, key assumptions, and limitations of the model are also well presented, and my understanding is that the authors share their model upon request. Hence, it is transparent. On the other hand, PRIME is a static model. The element of time is absent from the model, both in terms of lag times between exposures and diseases, and regarding risk factor exposure and mortality trends. Moreover, PRIME only models the impact of an intervention on disease mortality and is agnostic about disease incidence and prevalence, which makes its outputs less informative. Therefore, I argue that IMPACT$_{NCD}$ is more comprehensive than PRIME.

The second model is a proportional multi-state life table model that has been developed for the Australian and New Zealand populations.[253, 441] This is essentially a macrosimulation, but unlike a Markov macrosimulation, it allows for proportions of the population to be at multiple states simultaneously.[45] This model simulates multiple risk factors and the incidence, prevalence, and mortality of multiple diseases. It allows the exploration of the equity of the modelled interventions and it incorporates trends in disease incidence and mortality. Therefore, it seems reasonably reusable, but again it is not open source. The authors provide a technical appendix that describes the inputs, logic, key assumptions, and limitations of the model but they do not offer the actual model (in this case an Excel spreadsheet). Hence, it is reasonably transparent. On the other hand, this model does not model lag times between exposure and disease. It also models a closed cohort population which makes the model outputs not immediately interpretable to the real population.

So, does the adoption of a dynamic stochastic microsimulation approach for IMPACT$_{NCD}$ provides any benefits compared to a proportional multi-state life table model or it just increases complexity unnecessarily? Three main advantages stem naturally from a microsimulation that would have been hard to incorporate with any other approach. The first is that this approach allows for the heterogeneity of the population to be included in the uncertainty analysis (please refer to section 2.6 on page 62). The second is that it allows modelling explicitly the distributional nature of each risk factor exposure and the correlations between the exposure distributions. Therefore, high-risk individuals with multiple risk factors are naturally included in the simulation, and the 'differential vulnerability' pathway of the Diderichsen social production of disease model (section 1.4.2 on page 26) is modelled consequently. Finally, simulating individuals rather than cohorts dramatically increases the spectrum of the policies and interventions and the depth of their definition that can be simulated. All high-risk, targeted preventive interventions require an eligible, based on some predefined characteristics, population to be screened and then high-risk individuals (based again on predefined criteria) to be treated(section 1.5.1 on page 31). In a microsimulation setting defining eligibility and treatment criteria and alter them in a sensitivity analysis is easy, and once more, naturally stems from the structure of the model. Even, when modelling population-wide interventions that have agentic elements,

---

45 I. e. a traditional Markov macrosimulation would require a different state for each combination of diseases to allow for proportions of the population to have more than one diseases simultaneously. A proportional multi-state life table model eases this limitation and reduces the structural complexity of the model.

microsimulation offers the flexibility to explicitly model the differential penetration and effectiveness of the intervention based on individual characteristics (i. e. education, literacy, health attitudes, etc.). 'Hybrid' modelling approaches have been applied in the past. For example, the modelled intervention, disease incidence and case fatality may be modelled using microsimulation and disease prevalence and mortality may be modelled using proportional multi-state life table.[442]. These hybrid approaches provide further evidence that microsimulation has inherent benefits that are hard to be reproduced with a different modelling methodology. The microsimulation approach of IMPACT$_{NCD}$ was dictated from the wide spectrum of policies that had to be modelled to meet the aims and the objectives of these project (please refer to section 1.7 on page 40). The increased complexity, data and computational requirements of IMPACT$_{NCD}$ may be further rationalised in the future if IMPACT$_{NCD}$ get reused in other projects.

8.2.2 *Overarching themes*

In the following paragraphs I will focus on three recurrent themes in my results. The first theme is that structural, population-wide approaches were consistently more powerful than agentic, high-risk ones for primary prevention. In chapter 4, statins could only explain about a third of the observed cholesterol decline. In chapter 5, universal screening was three times less effective than a combination of population-wide structural policies. In chapters 6 and 7, the addition of more structural elements to existing policies for excess salt, and tobacco control greatly improved their effectiveness. This is in concordance with other modelling and empirical studies.[131–136]

This finding is particularly important because it strongly suggests that despite the observed concentration of risk in specific sub-populations, population-wide prevention remains more effective than high-risk approaches. The original concept by Geoffrey Rose remains topical now and for the foreseeable future, as recent risk factors trends continue (please refer to section 1.5.1 on page 31). In fact, as the prevalence of high-risk individuals in the future is predicted to decrease (figure 5.1 on page 126), identifying them through screening programmes will be increasingly inefficient, making population level policies a more attractive option.

The second overarching theme of my results is that generally, the equity of an intervention increases as its structural components increase. For instance, in chapters 6 and 7 the addition of more structural elements to existing policies for excess salt, and tobacco control improved their equity. This is consistent with existing theoretical and empirical evidence (please refer to section 1.5.2 on page 31).[112, 143–147]

However, the agentic policies that I modelled did not generate substantial inequalities, despite the theoretical arguments to the contrary (please refer to section 1.5.2 on page 31 and section 1.5.4 on page 33).[112, 143–147] In chapter 4, statins utilisation increased with deprivation for both sexes; subsequently, statins contribution to declining cholesterol also increased with deprivation for women. Similarly, in chapter 5 universal screening for CVD may not generate dramatic inequalities. The estimated equity of the current UK salt strategy in chapter 6 was more complex. Current salt policy, which has strong agentic

elements (media campaigns, food labelling, enhanced voluntary reformulation) had probably no effect on CVD inequalities; although, it is likely to generate inequalities for gastric cancer.

One explanation for the inconsistency between the theoretical expectations and the modelled equity estimates of agentic policies may be that planners and practitioners anticipated for these inequities and designed the intervention accordingly. Specifically, agentic policies require individuals to get exposed to the policy, react to the policy through behavioural change, and sustain the new behaviour.[146] Socioeconomic gradients may appear in any of these stages, which would generate inequalities. It is likely that with proper design and implementation of agentic interventions these socioeconomic gradients can be prevented. Using NHS Health Checks as an example, some areas may have allocated more resources in recruiting participants with a lower socioeconomic background. Then, participants with lower socioeconomic status are more likely to be offered preventive medication compared to similar individuals with a higher socioeconomic status, because deprivation is recognised as a contributing factor for CVD. And finally, the cost of medication is covered in full for the most socioeconomically deprived. Implementing, the same policy without these provisions could have deleterious effects in socioeconomic health inequalities.

Another plausible explanation for the gap between theoretical and simulated results could be that agentic policies generate health inequalities, which are not accounted for by the use of the Index of Multiple Deprivation; consequently, they are not reflected in the model. For example, agentic policies may discriminate against individuals with mental health issues, or inadequate health literacy;[356, 357] however, the published quantitative data do not allow for a modelling exploration of these dimensions of health inequalities. In any case, it seems that for agentic policies implementation details matter. While the risk of intervention-generated inequalities exists, provisions during the design of agentic interventions may counteract the anticipated inequity, at least partly.

Finally, the third theme that emerged from my results is the notion of joint prevention. Beyond the obvious and expected yet, still rarely explored joint prevention of NCDs,[135, 157] my results indicate that it may be feasible to jointly prevent NCDs and socioeconomic inequalities in health. Population-wide structural interventions targeting unhealthy diet, and/or smoking can potentially prevent multiple NCDs and tackle socioeconomic inequalities. This was highlighted in chapter 5 with the population-wide intervention scenario, and in chapters 6 and 7 with the structural salt and tobacco control scenarios. In fact, combining structural policies with carefully designed high-risk interventions targeting the most deprived groups may have even greater impact on inequalities and achieve the optimal effectiveness and equity (i. e. the combined scenario in chapter 5). However, the cost-effectiveness of these combined approaches remains unclear.

In conclusion, my results suggest that structural population-wide preventive policies are the cornerstone for effective and equitable joint primary prevention of NCDs and health inequalities. Interestingly, high-risk interventions targeting the most deprived may be useful in supporting the equity of structural policies. Nevertheless, this 'proportionate uni-

versalism' requires careful planning and implementation to avoid intervention-generated inequalities.

8.3   WHY MODEL? THE ROLE OF MODELLING IN PUBLIC HEALTH

In the previous paragraphs I summarised the key findings of my thesis for England.[46] I used modelling to quantify the impact of existing policies (i. e. statin prescription for primary prevention, universal screening for CVD, and the salt reduction strategy). Experimental methods for the evaluation of these national policies would require significant resources and complex stepped-wedge designs that are impractical to implement when there is substantial lag time between exposure and disease. Therefore, in these cases modelling may be the only realistic solution.

While alternatives exist for the evaluation of applied policies, modelling is the only methodology that can provide planners and policy makers with estimates regarding the impact of a policy on the population, before the actual implementation of the policy. Hence, multiple policies and variations of the policies can be simulated and the probability of success can be estimated against prespecified decision rules, allowing for evidence based decisions. In addition, any potential pitfalls can be identified during the simulation and the design of the policies can be further improved before they applied to the real population. After the implementation traditional evaluation methods can focus on the dimensions of the policy that were the most problematic or uncertain during the simulation. Modelling is considered a mature method for policy evaluation in other fields, but less so in public health.[443–447]

Beyond the evaluation of policies, modelling may be the only method[47] to study interventions that jointly prevent NCDs, even with no resource constraints. The difference in risk reversibility lag times renders the implementation of experimental studies impossible on ethical ground. For example, let us consider a randomised control trial to study the effect on gastric cancer of an intervention that reduces salt consumption. Participants in the intervention arm of the study would have reduced mortality because of the favourable effect of reduced salt consumption on CVD. Because of the likely shorter lag time for CVD, this would have manifested earlier than any effect on gastric cancer that has likely longer lag time and would might well have resulted in early termination of the trial.[448]

Apart from the obvious uses for modelling that have been stated in the previous paragraphs, the benefits of modelling are much wider than merely its outputs. The full modelling process from the conception, to implementation, scenario building, calibration, and sensitivity analysis has multiple benefits. In an editorial in the Journal of Artificial Societies and Social Simulation, Epstein posited "sixteen reasons other than prediction to build models".[156] In the following paragraphs I show how the work of this thesis illuminates key aspects of these sixteen benefits.

---

46 These findings could have been fundamentally different if I had simulated a different population. For example, the estimated effectiveness and equity of cardiovascular screening may have been very different if they had been implemented in a sub Saharan African country.

47 The alternative is relevant natural experiments.

EXPLAIN: Offering plausible explanation or illustration of a phenomenon is perhaps the greatest advantage of modelling. For instance, IMPACT$_{NCD}$ can help to illustrate how the interplay of exposure to risk factors with socioeconomic circumstances creates the socioeconomic patterns in CVD, lung, and gastric cancers morbidity and mortality. As Epstein argues, explain is much different from predict.[156] We can explain with great detail how blood clots are forming in the atherosclerotic coronary arteries and cause AMIs; however, we cannot predict accurately when this will happen. Explaining a phenomenon may be more important for public health policy because public health policy makers are only rarely experts in public health or epidemiology. Therefore, models can be the means of effective communication between experts and policy makers.

GUIDE DATA COLLECTION: Many believe that models are built to account for the data. This is often the case; however, the most useful models are those which precede the data and guide data collection. For instance, the existence of Higgs boson was indicated in the 1960s from the Standard Model of physics. Since then, physicists have been searching for it for more than 50 years and they only managed to confirm its existence in 2013.[449] Unfortunately, similar high profile examples are still rare in public health modelling. Fortunately, IMPACT$_{NCD}$ might have provided an illustrative example. In section 1.3 on page 17 it was obvious that empirical evidence regarding salt risk reversibility for gastric cancer was lacking. However, IMPACT$_{NCD}$ validation assuming the existence of salt risk reversibility with a mean lag time of 8 years performed well both for gastric cancer incidence and mortality. Hence, my results may suggest that the risk of gastric cancer from exposure to excess salt is fully reversible within about eight years from a successful reduction of salt consumption.[48]

ILLUMINATE CORE DYNAMICS: In chapter 1, I described how the population dynamics, trends in risk factor exposures, and lag times between exposure and disease drive the trends in morbidity and mortality. In IMPACT$_{NCD}$ I meticulously tried to capture these dynamics because they are particularly important in explaining morbidity and mortality in the population. Traditional epidemiological methods usually fail to describe and quantify these dynamics in a form useful for decision-making. For example, traditional methods in modelling may quantify population ageing, the increasing trend in obesity prevalence and the decreasing trends in SBP and cholesterol. However, in isolation those findings cannot pinpoint whether CVD incidence will increase or decrease in the population. Modelling can integrate this information and produce the desired output. In chapter 7, I also used this feature of modelling to describe the dynamics of two tobacco endgame policies.

SUGGEST DYNAMICAL ANALOGIES: Models are abstracted idealisations of reality; hence, analogies between phenomena are more easily identified when these phenomena are modelled. For instance, the spread of the obesity epidemic over the past few decades

---

48 Although because IMPACT$_{NCD}$ does not account for a declining trend in the prevalence of Helicobacter pylori infection, a known gastric cancer risk factor that is also associated with high salt diet, other plausible explanations also exist.[450, 451]

has similarities with the spread of communicable diseases and both can be described by similar mathematical equations.[452] The recognition of these analogies provides better insight and may open new avenues for exploration. In IMPACT$_{NCD}$ I showed that the incidence of multiple NCDs and their trends can be modelled using the same modelling framework, suggesting that despite their diversity they share some common elements. I did the same for the different policies.

DISCOVER NEW QUESTIONS: As models integrate all the relevant information to simulate a phenomenon or a system and answer old questions, new questions are discovered. In my thesis these new questions may be: what is the most effective and equitable combination of structural and targeted policies to prevent CVD? or why did the modelled agentic policies not generate substantial inequalities?

PROMOTE A SCIENTIFIC HABIT OF MIND: This is perhaps the most important of all because it can potentially benefit public health practitioners and policy makers. The process of model building is a journey of scientific enquiry. Information from diverse sources needs to be collected, challenged, and then combined in the model. In this journey practitioners, planners, and policy makers may guide or lead the modelling team because usually they have a better insight into the complex system to be modelled. Therefore, they know what components of the system are less important, and what assumptions are reasonable. Through the modelling process though, practitioners, planners, and policy makers may explicitly formulate their insights and share them with others; hence, these now explicit insights and beliefs can be scientifically challenged and scrutinised before they inform the model. During the design and application of IMPACT$_{NCD}$ I followed these principles.

BOUND OUTCOMES TO PLAUSIBLE RANGES: I find plausible ranges more interesting and useful than point estimates, which are extremely unlikely to be accurate. The ideal scenario in chapter 6 and ideal minimum scenario in chapter 7 do exactly that. They set the upper boundary of cases that can be prevented or postponed from policies targeting salt and tobacco respectively. Comparative risk assessments provide the same information and their results are already used to inform policy-making. The advantage of more complex approaches is that they can provide the boundaries for more complex scenarios and estimate the dynamics of the boundaries over time.

ILLUMINATE CORE UNCERTAINTIES: Traditional epidemiological methods that are based on frequentist methods can only quantify the uncertainty arising from sampling. On the contrary, a more comprehensive approach to uncertainty is allowed with modelling (I provide a brief summary on modelling uncertainty in section 2.6 on page 62). In particular, uncertainty can be studied and quantified in models, but most importantly can be dissected to highlight core uncertainties that if improved can lead to more certain model outputs and better decisions. Hence, policy makers can decide to delay their decision, and allocate

resources to specific areas of research in order to reduce model uncertainty before they proceed to decision-making.

OFFER CRISIS OPTIONS IN NEAR REAL TIME: When models are designed to be re-usable, they can provide timely estimates when time is important. Such emergencies are unlikely for NCDs but are quite common in the field of infectious diseases and environmental health. Synthetic populations may be useful in these time critical situations because they can be used by a wide range of models for NCD, communicable disease, or environmental health modelling. The synthetic population of IMPACT$_{NCD}$ can also be used to a communicable disease model.

DEMONSTRATE TRADE-OFFS / SUGGEST EFFICIENCIES: This is another useful benefit of modelling in public health, especially in times of austerity and resource constraints. Modelling can be used to estimate the best mixture of policies to maximise health in the population while minimising inequalities and costs. Although IMPACT$_{NCD}$ cannot be used to estimate cost-effectiveness yet, the equity summary chart was designed to illustrate efficiencies and trade-offs in effectiveness and equity of the modelled policies (section 2.7.2.2 on page 66). Another example from my results is that in the future, screening strategies will be less effective because of the reduction of the high-risk prevalence pool as suggested in chapter 5 (figure 5.1 on page 126).

CHALLENGE THE ROBUSTNESS OF PREVAILING THEORY: I have to also add here prevailing theory-based practice. The superior effectiveness of population-wide interventions has been demonstrated multiple times.[131–136] Yet, high-risk, agentic approaches are still in the core of national prevention strategies. Statins for primary prevention and the NHS Health Checks programme are two examples. Models like IMPACT$_{NCD}$ can challenge current policies and provide leverage for their improvement or replacement. Interestingly, another model may explain why policy makers opt for more agentic approaches and "why cure crowds out prevention", as the higher revenue of the curative sector is used to influence policy.[453]

EXPOSE PREVAILING WISDOM AS INCOMPATIBLE WITH AVAILABLE DATA: A great example of this is in chapter 4. Many epidemiologists and clinicians maintain that statins were the main driver of cholesterol reduction observed in many countries over the past two decades, including England. In chapter 4, I provided evidence that this is not supported by the data, as the observed reduction was much steeper than would be expected given statins utilisation and their effect size. A phenomenon that it has been observed in other countries also.[309–312]

TRAIN PRACTITIONERS: Practitioners may be trained not only through participation in model building but also through simple interaction with the model. For instance, by inputting scenario parameters and observing the numerical and graphical model outputs, practitioners may develop a more comprehensive understanding of how and to what ex-

tent specific policy elements can influence population health. IMPACT$_{NCD}$ findings have already been published in peer-reviewed journals, and have been presented in national conferences and local meetings, where public health trainees were also attending.[298–301, 454] Nevertheless, my aspiration is to expand the current user interface of IMPACT$_{NCD}$ and allow users to directly interact with the model for a better learning experience.

DISCIPLINE THE POLICY DIALOGUE: Actually, policy dialogue can be evolved, facilitated, and progressed around a modelling exercise. All the aforementioned benefits of modelling can then provide insight, enrich, and discipline the discussion. As I have previously mentioned, many of my findings have already been published, and presented in conferences and meetings. Therefore, they are already part of the policy dialogue. Additionally, I had the honour to present and discuss my findings directly to national policy-making committees. Finally, I already work with Liverpool Local Authorities to create a local model and I hope that this will facilitate and progress the policy dialogue locally.

REVEAL THE APPARENTLY SIMPLE (COMPLEX) TO BE COMPLEX (SIMPLE):   Lastly, the promotion of scientific enquiry and explicit assumptions that modelling offers can sometimes reveal that 'simplicity' and 'complexity' are not absolute terms but relative and depend on the context. I will use two examples from my personal experience to further clarify this. Before this research project I had mistakenly believed that identifying preventive policies that are both effective and equitable would be a complex task. After building and experimenting with IMPACT$_{NCD}$, the same task seems much simpler now. On the other hand, during my literature review I got the impression that high-risk policies were synonymous with intervention-generated inequalities. Now I understand that this is a far more complex issue.

EDUCATE THE GENERAL PUBLIC: I have to add here also 'involve the general public'. My results suggest that structural policies are key in the primary prevention of NCDs. However, implementation of such policies usually requires strong public support and involvement. Model outputs can be used to inform powerful media campaigns in order to explain structural changes and gain public support and involvement for radical policies.

8.4   IMPLICATIONS FOR PLANNERS POLICY MAKERS AND CLINICIANS

My thesis was mainly written to inform policy makers and planners. Isolated implications of my findings have been described separately, in each result chapter. Here, I will reiterate two fundamental arguments stemming from my thesis. First, primary prevention policies need to have strong structural components in order to maximise effectiveness and have an impact on health inequalities. Current preventive strategies can be further optimised by the addition of more structural policies. Furthermore, new radical approaches may be needed to substantially reduce the burden of NCDs and tackle socioeconomic health inequalities. Second, public health modelling is mature enough to support decision-making

in public health and guide policy. Therefore, new partnerships and collaborations are necessary between the modelling, policy and decision-making communities.

My findings, support the principle that primary prevention needs to be moved away from health care interventions and towards more structural approaches for increased effectiveness and equity. Of course, this does not diminish the importance of opportunistic screening and medical consultation in health care settings for individuals. Nor does it downgrade clinical judgement. However, on the population level these interventions are just a drop in the ocean, despite their importance on the individual level.

8.5 LIMITATIONS

In chapter 2, I summarised the main limitations and assumptions of IMPACT$_{NCD}$ (table 2.1 on page 69). Here I will discuss them in depth, along with their possible direction and magnitude of the introduced bias. I will also discuss the less technical but perhaps more important scenario assumptions and limitations that were not included in this table.

In line with the IMPACT$_{NCD}$ description I will start with the assumptions and limitations of the population module. First, the demographic engine ignores migration flows. This affects the synthetic population growth over time, as well as the distribution of the risk factors in the population. Long term forecasting of migration flows is difficult; however, assuming that the net migration[49] will continue its increasing positive trend,[455] IMPACT$_{NCD}$ underestimates the synthetic population projections by about 0.4 % every year. As a result, the estimated numbers of cases and deaths prevented or postponed are likely to be underestimated also. Unfortunately, forecasting the absolute number of immigrants and emigrants is the simpler part of the problem. In the context of microsimulation it is necessary to know migrants' individual characteristics; from their age and sex to their behavioural and biological risk factors. This level of information is lacking on an individual level and is patchy on an aggregated level because of the diversity of immigrant populations. What can be safely assumed though, is that immigrants face more barriers in accessing the health care system than natives.[456] Hence, the exclusion of immigrant flows from IMPACT$_{NCD}$ is likely to overestimate the effect and equity of high-risk, agentic policies and the opposite is likely for the structural ones.

The second limitation of the population module is the rather simplistic approach to modelling the dynamics of deprivation. The first issue with this, is that the Index of Multiple Deprivation is a marker of relative deprivation that is updated every few years to take into account new sources of information. I assumed that all versions of the Index of Multiple Deprivation are the same, ignoring the likely overall absolute improvements of the socioeconomic conditions in England between 2001 and 2012. The bias from this is unclear because the level and speed of the assumed improvement in each quintile is poorly quantified. The second issue is that I used the QIMD as the only marker of socioeconomic deprivation in the dynamic model. While the initial synthetic population contains also the household income and the employment status of the head of the household (section 2.3.2.1

---

49 Defined as immigration – emigration.

on page 46), this information is not considered in the equations that define the behavioural and biological risk factors of the synthetic individuals. I chose this approach for simplicity; however, I plan in the future to experiment with nested hierarchical models that include individual and household markers of deprivation, nested within the area level deprivation marker. The third issue with the deprivation dynamics in the model is that social mobility is not considered. Synthetic individuals are permanently allocated into one of the five socioeconomic groups for the full course of the simulation. This prohibits IMPACT$_{NCD}$ to model the positive feedback loop in the Diderichsen model (please refer to section 1.4.2 on page 26). Technically, it would be easy to simulate social mobility in IMPACT$_{NCD}$ if the impact of social mobility on the behavioural risk factors was known. Unfortunately, it is not clear to what extent individuals that improve or worsen their socioeconomic position during their life course adopt healthier or less healthy behaviours, respectively. Because these assumptions and simplifications apply to all scenarios, it is unlikely that they may have seriously biased the equity ranking of the policies.

The third limitation of the population module is that the exposures in the synthetic population are informed by one source (HSE). Hence, any biases in the HSE are projected to the synthetic population. This may introduce bias mainly in two possible ways. First, through selection bias. HSE adjusts for selection bias arising from its sampling since 2003, however the main assumption of this adjustment is that non-responders have similar characteristics to the responders. Given the socioeconomic patterns of most risk factor exposures, it is likely that the socioeconomic gradients of exposures in the synthetic population are underestimated. The second possible source of bias is that physical activity and fruit and vegetable consumption were self-reported. The questionnaires that have been used over the years for the measurement of physical activity were not identical, and responses to the same questions may have changed culturally over time, for both physical activity and fruit and vegetable consumption. Therefore, the extraction of temporal trends for these two variables may be problematic. Most importantly, the self-reported physical activity was poorly correlated with accelerometer measurements, raising questions regarding the validity of this exposure measurement and indicating social desirability bias.[457] This is the main reason I have avoided modelling policies involving physical activity. The fruit and vegetable consumption was recorded using 24 hour recall in HSE. The point estimates from the HSE were lower than those of NDNS that used 4 day diary records, by slightly less than a portion. However, the validation of NDNS suggested under-reporting of overall calories intake and possibly social desirability bias (please also refer to section 1.3.3.2 on page 19).[54] Hence, I decided to use HSE estimates. Additionally, the one-off measurement of fruit and vegetable consumption may overestimate the variance of the 'usual exposure' distribution of the population, because of the within person variation.[32] IMPACT$_{NCD}$ adjusts for that by allocating different synthetic individuals to the extremes of the consumption distribution for every simulated year, essentially diluting the bias to the whole synthetic population at the expense of increased uncertainty.

The disease module has a different set of assumptions and limitations. The multiplicative risks assumption that IMPACT$_{NCD}$ uses is the norm for comparative risk assessments and modelling.[30, 32, 172, 458] Specifically for CVD this assumption is supported by empir-

ical studies.[103, 459] However, evidence is lacking regarding cancers. The case is similar for the 100 % risk reversibility assumption. The only case where evidence disputes this assumption is for SBP and CHD, for which risk reversibility is about 75 % (please refer to section 1.3.7.1 on page 22). Therefore, IMPACT$_{NCD}$ is likely to overestimate the effect of interventions targeting SBP. Another assumption in IMPACT$_{NCD}$ is that the exposure to risk factors has the same effect on disease incidence and mortality. This assumption was inherited from some of the meta-analyses used in IMPACT$_{NCD}$ that included indistinguishably both CVD incidence and CVD mortality primary studies to extract the effect size of BMI and diabetes mellitus.[74] Similarly, I used effect sizes from mortality studies to model incidence.[78, 81] Consequently, I assumed that exposure to risk factors has no effect on case fatality. For instance, the long term survival after an AMI in the model is the same irrespective of whether the patient quits smoking after the event or not. In fact, the risk may almost halve.[460] Since I have only modelled primary prevention policies, these assumptions have no effect on the estimated cases prevented or postponed and may underestimate the deaths prevented or postponed.

Finally, with the exception of salt and SBP, I have extracted the magnitude[50] of the associations between behavioural and biological risk factors from the HSE, to allow for greater granularity. Because of the cross-sectional design of HSE, reverse causality may have biased these estimates, attenuating the associations. Hence, this may have underestimated the effect of population-wide interventions.

Overall, the direction of bias from these technical assumptions and limitations is towards underestimation of the cases prevented or postponed that were estimated by IMPACT$_{NCD}$. The underestimation is likely to be greater for the structural interventions, hence it further supports the argument that structural policies are more effective than high-risk, agentic ones. The validation of the model suggests that the introduced bias of all the assumptions in the model is small.

In summary, the IMPACT$_{NCD}$ framework is likely to be balanced in translating changes in risk factor exposures into changes in NCDs incidence and mortality, with overestimates balanced by underestimates. This also allows for the scenario assumptions to be reflected in model outputs. Depending on the desired use of the model, this property of IMPACT$_{NCD}$ can be seen as a limitation or a strength. If the purpose of modelling is forecasting, then this is probably a limitation because model outputs are sensitive to the input assumptions. However, if the purpose of modelling is to support policy makers in their decision process, then the sensitivity of the model is a great strength. Due to the comprehensiveness of IMPACT$_{NCD}$, it requires detailed policy specification for the modelled scenarios. Consequently, it encourages the user to describe in detail the policy, and when information is lacking, to make explicit assumptions about who will get exposed to the policy, who will engage and respond, and in what way.[51] The way a policy is specified will have a great impact on its estimated effectiveness and equity. The explicit assumptions though, can be

---

50 As a reminder, the existence and the direction of the associations were informed from published longitudinal studies.

51 It is worth noting here that simpler modelling approaches may require less specific scenarios. The price is implicit assumptions that cannot be contested, and reduced transparency.

contested, further explored in sensitivity analysis, or their uncertainty propagated to the outputs. Essentially, this allows IMPACT$_{NCD}$ to realise the advantages of modelling that have been described in section 8.3 on page 168.

8.6   FUTURE PLANS AND CHALLENGES

In this section I will first discuss the immediate plans for developing IMPACT$_{NCD}$ further. Then, I will summarise my vision for public health modelling in the next decade and some of the challenges that the discipline may face.

In my thesis I have described the IMPACT$_{NCD}$ framework, which is a generic framework for NCD modelling, and I have showcased an implementation of the framework as a proof of concept. The immediate next step is to extend the current implementation in five axes. The first is to include more risk factors in the simulation; alcohol and other markers of poor diet such as fibre, sugars, and fats consumption are the most obvious ones. The second axis is to expand the modelled diseases; more cancers, chronic obstructive pulmonary disease, and dementia are candidates for this.

The third axis is more ambitious. I would like to create a separate policy layer framework to systematically model preventive policies and interventions. User assumptions about the policies will perhaps be vital for this, and a user interface will be necessary to help the user interact directly with the model and test these assumption in real time. The creation of a registry of preventive policies that will contain empirical data regarding policy aspects, and reasonable assumptions about these policies may follow, to be used as template policies for IMPACT$_{NCD}$.

The fourth axis is to expand the available modules and therefore, the usability and re-usability of the model. The obvious step is to include a health care intervention module after the population and disease modules. This will allow scenarios involving secondary prevention. Then, perhaps a health economics module to estimate cost-effectiveness along with effectiveness and equity of the policies. Another module could back-project the simulation in order to explain past exposure and disease trends. The addition of an 'early exposures' module is a longer term plan to allow scenarios of childhood prevention and facilitate 'life course' analyses.

The final expansion axis is to include more populations in the model. Scotland and the US can be among the first new countries to be modelled, as both have a series of health surveys that are freely available to researchers. A less obvious opportunity is to extend the model to local populations in smaller geographical areas; i. e. model populations within cities. This is a field largely unexplored and the interventions that can be implemented at city level are much different from the national policy options. In England, many public health responsibilities have been moved from the central government, down to local authorities, making a local implementation of IMPACT$_{NCD}$ very topical.

8.6.1 *Public health modelling in 2030*

The benefits of modelling for public health that I listed in section 8.3 on page 168 could be realised immediately, at least in part, with the modelling framework reported in this thesis. However, new opportunities and barriers lie ahead. The first part of the 21st century will be probably known to the future generations as the beginning of a new era, the 'age of Big Data'.[461] A huge amount of information is collected everyday and can be potentially linked to create new knowledge. Data science, predictive analytics, machine learning, and artificial intelligence are rapidly evolving scientific fields that are driven by the eruption of Big Data. The use of big analytics in Big Data in health, despite the 'teething' problems, might potentially revolutionise health care.[462] In this new data rich era, the accuracy of existing clinical prediction models may substantially increase. Furthermore, the necessary resources to calibrate them for new populations and expand them to new diseases may be dramatically decreased.[463]

What might be the impact of Big Data in future public health policy and public health modelling? Imagine a national registry of the English population that links health care records, with information about consumption, and information about behavioural characteristics collected from social media. Buchan *et al.* argued that the aggregation of health data from multiple, diverse sources about an individual will lead to the creation of a 'health avatar', "…the electronic representation of an individual's health as directly measured or inferred by statistical models or clinicians".[464] A public health model could have access to the health avatars of a population and identify high-risk individuals in the population for certain diseases. Then, the model could wait for the right moment to invite them for a consultation with a health care professional in order to maximise response; i. e. a few days after their favourite artist had an AMI or after setting up a personalised awareness campaign in their newsfeed. The model could monitor whether the avatars indicating that the individuals had altered their behaviour towards healthier options and perhaps intervene if not. Another model could raise public awareness and support for a structural intervention using social marketing approaches, and notify policy makers when there would be enough support in the population to implement the change by monitoring social media. In a less futuristic scenario, this wealth of information may be used to better define the effectiveness and equity of the applied preventive policies. Simulation modelling can be used to generate independent estimates for comparison with the collected data. Any convergences or divergences could be explored and used for better calibration of the simulation model, or flag potential issues in the collected data. Big Data may allow the development of 'precision public health' (an analogy to precision medicine), in which public health modelling is a vital component.[465]

A synthetic population (of avatars) at some point may become more credible than a population averaged model; but may not necessarily be more useful for supporting public health policy decisions. Leaving aside ethical considerations, Big Data does not necessarily equate population representative data. Training predictive models in partial data reduces their predictive power and may lead to biased predictions. Therefore, using entirely data driven prediction models trained with Big Data, to predict about the population and inform

policy-making, is potentially dangerous. A vivid example of this was how the Bayesian age period cohort model I used in the validation (section 3.4.3 on page 98) got 'trapped' in an artefact of the data. In contrast, public health simulation models that aim to explain rather than predict, and integrate information from multiple sources including traditional research methods and experts' opinion without over relaying on Big Data, may be proved more useful to policy makers. These models may also be used to augment Big Data by adding information to sparse areas of the data in an attempt to improve the robustness of predictive models.[466]

8.7   PERSONAL REFLECTIONS

As a consequence of my interest in public health, I decided to embark on this PhD journey with no previous experience in modelling, but with a strong desire to learn and further develop. Coming from a clinical background and with strong foundations in statistics and computer programming, public health modelling seemed like the discipline that would combine my skills and interests. I was not mistaken. During these three years of my research, modelling became the vehicle for me to widen my understanding in public health, policy, and epidemiology. Most importantly though, it enabled me to communicate my ideas, and discuss them with a wide range of people; from data scientists and statisticians, to public health practitioners, policy makers, and the general public. This enabled me to improve my communication skills outside the clinical setting, and to realise that good ideas need even better support and communication to flourish. After all, Voltaire was probably right that 'common sense is not that common'.[467] Finally, I am proud because unlike many other PhD projects that stop after three years, I will continue to develop and expand IMPACT$_{NCD}$ for the foreseeable future and I hope that other members of the open-source community will eventually get involved with it.

9

CONCLUSIONS

This initial journey through the fields of epidemiology, public health policy, and computational statistics has almost finished. I have described the great and unequal burden of NCDs in England and the preventive policy typologies that can be used to guide policy-making. I have also described how simulation modelling can support policy makers in their decision-making process, by integrating all the available information and providing insights that take into account the complex dynamics of exposures and diseases in the population.

I will end by summarising the three key messages that emerged from my research. First, although existing primary prevention policies for NCDs appeared to be effective overall, the addition of structural elements to these policies may further optimise their effectiveness and equity. Second, high-risk interventions that target the most deprived groups combined with structural policies, may achieve the best impact on reducing disease burden and health inequalities. Finally, modelling offers a potentially powerful approach to assess the impact of existing policies when more traditional methods are impractical. More importantly, simulation modelling is essential for the design of new fit-for-purpose policies that will take into account the complex nature and dynamics of the specific population that they are designed for.

Part III

APPENDICES

METHODS APPENDIX

A.1 SALT STOCHASTIC PROCESS

STAGE 1   The sodium surveys report several percentiles of the 24 h urine sodium distribution by age group and sex. I used least squares estimation to fit known continuous univariate distributions[52] to the reported percentiles. The distribution with the best fit was selected and used for further calculations. To avoid bias from possible outliers to the extremes of the reported percentiles, I used the formula $weights = 1/(|0.5 - percentile| + 1)$ to give higher weights to percentiles around the median. The R package 'rriskDistributions' was used in this stage.[468] The end result of this stage was that for each age group, sex, and year of sodium survey I estimated a continuous distribution of sodium excretion. For instance, a triangular distribution was selected for men, aged 19 – 24 in 2001 with parameters min ≈ 5.18, mode ≈ 7.3, and max ≈ 21.07 (figure A.1 on the following page).

STAGE 2   The four sodium surveys were performed in years 2001, 2006, 2008, and 2011. I used the nearest year HSE that individual level data from spot urine sodium was available and I converted the spot urine sodium to 24 h sodium, using the INTERSALT equation for Northern Europe.[219, 239] For each HSE participant different coefficients of the INTERSALT equation were sampled from the normal distributions with mean equal to the reported coefficient and SD equal to the reported SE of the respective coefficient. Finally, 24 h sodium (in mEq/day) was converted to salt (g/day) using the formula $1 mEq$ of sodium $= 58.5 * 10^{-3} g$ of salt.

STAGE 3   In this stage, the percentile rank[53] of the estimated salt consumption for each HSE participant was calculated by 5-year age group, sex, and year. Then, the estimated salt consumption values from stage 2 are substituted by equal number of values that were drawn from the respective (by age group, sex, and year) salt distribution that was estimated in stage 1, based on the equality of percentile ranks. For example, consider a participant whose salt consumption was estimated in stage 2, at 10 g/d. Let us suppose that the percentile rank for his/her respective age group, sex and year corresponds to 0.6. Then in this step, a set of numbers will be drawn from the respective distribution estimated in stage 1 and the value with percentile rank of 0.6 will replace the 10 g/day salt consumption. Therefore, by the end of this stage, the individual level data from HSE years 2003, 2006, 2009, and

---

52 Normal, beta, Cauchy, logistic, t, chi square, non-central chi square, exponential, F, gamma, log-normal, Weibull, triangular, PERT, truncated normal and Gompertz.

53 For the percentile rank the formula $R_{textpercentile} = (R - 1)/(n - 1)$ was used, where $R_{percentile}$ is the percentile rank and $R = (R_1, \dots, R_n)$ is the rank vector constructed from a random observation vector $(X_1, \dots, X_n)$.
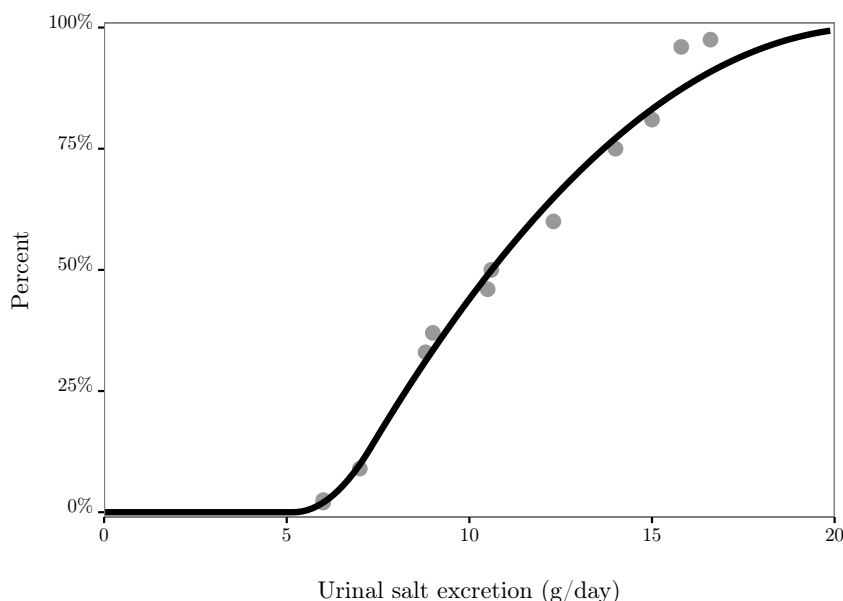
Figure A.1: Plot of the cumulative distribution function of the best fit distribution (curve) against known quantiles (points) for men, aged 19 – 24. Data from sodium survey 2001.[242]

2012 regarding salt consumption, have very similar statistical properties as those reported in sodium surveys.

STAGE 4   Quantile regression models were fitted to the series of HSE data with salt consumption from the previous step as the dependent variable, and ln(year of the survey −1997), third-degree orthogonal polynomial of age, sex, QIMD and their first order interaction as the independent variables. The models were fitted for the 0.01, 0.05, 0.10, 0.15, . . . , 0.90, 0.95, 0.99 percentiles.

STAGE 5   Stages 2 to 4 were repeated 1000 times and 1000 quantile regression models were built.[54]

STAGE 6   During the simulation, the percentile rank of salt consumption for each synthetic individual in IMPACT$_{NCD}$ is calculated from the previous year salt consumption stratified by age, sex, and QIMD. A quantile regression model is drawn from the models estimated in stage 5 and is used to predict the respective percentiles of the salt distribution by age, sex, QIMD and year. Then, for each synthetic individual a minimum and a maximum value for salt consumption is defined based on their percentile rank. For example, if the percentile rank of a synthetic individual is 0.23 the minimum and maximum values will be predicted from the 0.20 and 0.25 percentile regression models respectively. Finally, a new salt consumption for current simulation year will be drawn from the uniform distri-

54   The model is available at https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/Lagtimes/salt.rq.rda and the coefficients of all 1000 models at https://github.com/ChristK/IMPACTncd/blob/Thesis_model_version/Lagtimes/salt.rq.coef.rda.

bution with minimum and maximum values those that were predicted from the 0.20 and 0.25 percentile regression models respectively.

A.2 EQUITY SUMMARY CHART

Consider again the simple example of a population that consists of only two mutually exclusive socioeconomic groups, the 'deprived' and the 'affluent', with different disease incidence in each group. Then the disease incident cases of the whole population $I = I_{deprived} + I_{affluent}$, where $I_{deprived}$ is the cases in the 'deprived' group and $I_{affluent}$ is the cases in the 'affluent' group. Also by definition, absolute inequality $E_{abs} = I_{deprived} - I_{affluent}$, and relative inequality $E_{rel} = I_{deprived}/I_{affluent}$.

For a given overall reduction in disease incident cases across groups (ΔI), the distribution of the reduction among the two groups can be described as $I_{deprived} - a$ and $I_{affluent} - b$, where $a + b = \Delta I$. The post-intervention absolute and relative inequality is $E'_{abs} = I_{deprived} - I_{affluent} - a + b$ and $E'_{rel} = (I_{deprived} - a)/(I_{affluent} - b)$, respectively. Assuming that the intervention has approximately no effect on the size of the two subgroups, for $E_{rel} = E_{rel}el'$ it can be shown that $a = E_{rel} * b$ and $E_{abs} - E'_{abs} = \Delta I * (E_{rel} - 1)/(E_{rel} + 1)$, which on the equity summary chart is a line that represents the 'equity line' for this population (Figure 12). Interventions above the equity line decrease relative socioeconomic inequalities and interventions below the line increase it. Moreover, the vertical distance from the equity line is proportional to the impact of the intervention on relative socioeconomic inequalities.

For the generalisation of the previous example to populations with more than two levels of socioeconomic deprivation and unequal sizes of the socioeconomic groups SII and RII have to be used. Hence, using the notation of the previous example $SII = I_{deprived} - I_{affluent}$ and $RII = I_{deprived}/I_{affluent}$, where this time $I_{deprived}$ and $I_{affluent}$ are extrapolated from the linear regression that was used for SII. Therefore, incident cases I for the whole population are not equal to $I_{deprived} + I_{affluent}$. From these formulas it can be shown that $I_{affluent} = SII/(RII - 1)$ and $I_{deprived} = RII * SII/(RII - 1)$. For the incident cases I of the whole population holds that $I = r * I_{deprived} + (1 - r) * I_{affluent}$, where $0 \leq r \leq 1$ and r depends on the distribution of inequality in the population. From the previous formulas it can be shown that $r = I/SII - 1/(RII - 1)$.

After the intervention, the new incidence $I' = r' * I'_{deprived} + (1 - r') * I'_{affluent}$ and will result in a new SII' and RII'. Because r' is dependent on the impact of intervention on the different socioeconomic groups, the equity line cannot be defined as in the previous simplified example. However for a given intervention, r' can be estimated and assuming there is an SII'' for $RII = RII'$. It can be shown that $SII'' = I' * (RII - 1)/(r' * (RII - 1) + 1)$ and SII'' increases monotonically as RII increases. Therefore, $SII - SII''$ decreases monotonically as RII increases. Because the horizontal axis of the equity summary chart represents $\Delta I = I - I'$ (cases prevented or postponed) and the vertical axis $\Delta SII = SII - SII'$, (impact on absolute socioeconomic health inequalities) of the modelled policies, the latter can be directly plotted on the chart. The point ($\Delta I, SII - SII''$) which represents the hypothetical policy impact on absolute socioeconomic health inequalities that would have caused
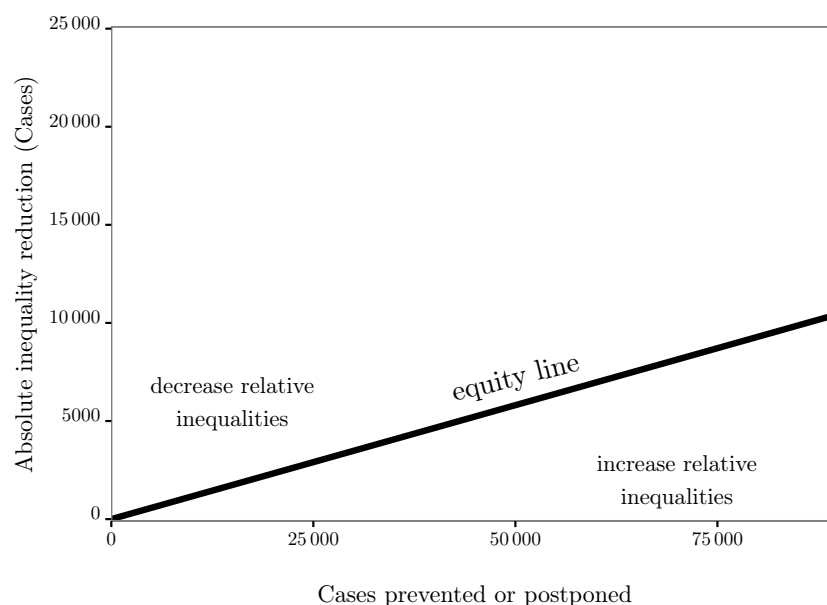
Figure A.2: Simplified equity summary chart assuming only two mutually exclusive socioeconomic groups and assuming that interventions do not alter population size of the groups.

a neutral effect on relative socioeconomic health inequalities for the same policy effectiveness can also be plotted. For multiple policies, a constrained β spline can be fitted to these points that would represent the 'equity curve' with properties approximately similar to the equity line. Modelled policies above the equity curve decrease relative socioeconomic health inequality and scenarios below the equity curve increase it. The vertical distance from the equity curve, approximates the impact of the scenario on relative inequality. Consequently, the health equity impact chart presents on a two axes chart, the impact of the intervention on disease incidence, absolute and relative inequality, in agreement with recommendations by health inequalities experts (section 1.4.4 on page 29).

Table A.1: IMPACT$_{NCD}$ distributions that were used as inputs for the simulations. Numbers are rounded.

Variable[source]	Sex	Age group	Distribution
Relative risks of relevant risk factors for CHD			
Active smoking[37]	Men	30 - 44	log-normal (mean = ln(5.51), SD = ln(12.3 / 5.51) / 1.96)
		45 - 59	log-normal (mean = ln(3.04), SD = ln(3.48 / 3.04) / 1.96)
		60 - 69	log-normal (mean = ln(1.88), SD = ln(2.08 / 1.88) / 1.96)
		70 - 79	log-normal (mean = ln(1.44), SD = ln(1.63 / 1.44) / 1.96)
	Women	30 - 44	log-normal (mean = ln(2.26), SD = ln(6.14 / 2.26) / 1.96)
		45 - 59	log-normal (mean = ln(3.78), SD = ln(4.62 / 3.78) / 1.96)
		60 - 69	log-normal (mean = ln(2.53), SD = ln(2.87 / 2.53) / 1.96)
		70 - 79	log-normal (mean = ln(1.68), SD = ln(1.93 / 1.68) / 1.96)
		80 - 84	log-normal (mean = ln(1.38), SD = ln(1.77 / 1.38) / 1.96)
Ex-smoking[279]	Men	30 - 84	log-normal (mean = ln(1.25), SD = ln(1.32 / 1.25) / 1.96)

…continued

Variable[source]	Sex	Age group	Distribution
	Women	30 - 84	log-normal (mean = ln(1.2), SD = ln(1.34 / 1.2) / 1.96)
Environmental tobacco smoking[280]	Both	30 - 84	log-normal (mean = ln(1.26), SD = ln(1.38 / 1.26) / 1.96)
SBP[78]	Men	30 - 49	log-normal (mean = ln(0.5), SD = ln(0.54 / 0.5) / 1.96)
		50 - 59	log-normal (mean = ln(0.5), SD = ln(0.52 / 0.5) / 1.96)
		60 - 69	log-normal (mean = ln(0.55), SD = ln(0.57 / 0.55) / 1.96)
		70 - 74	log-normal (mean = ln(0.62), SD = ln(0.64 / 0.62) / 1.96)
		80 - 84	log-normal (mean = ln(0.69), SD = ln(0.73 / 0.69) / 1.96)
	Women	30 - 49	log-normal (mean = ln(0.4), SD = ln(0.49 / 0.4) / 1.96)
		50 - 59	log-normal (mean = ln(0.49), SD = ln(0.54 / 0.49) / 1.96)
		60 - 69	log-normal (mean = ln(0.5), SD = ln(0.61 / 0.5) / 1.96)
		70 - 74	log-normal (mean = ln(0.55), SD = ln(0.58 / 0.55) / 1.96)

...continued

Variable[source]	Sex	Age group	Distribution
		80 - 84	log-normal (mean = ln(0.64), SD = ln(0.68 / 0.64) / 1.96)
Total cholesterol[81]	Both	30 - 49	log-normal (mean = ln(0.49), SD = ln(0.52 / 0.49) / 1.96)
		50 - 59	log-normal (mean = ln(0.62), SD = ln(0.65 / 0.62) / 1.96)
		60 - 69	log-normal (mean = ln(0.74), SD = ln(0.76 / 0.74) / 1.96)
		70 - 74	log-normal (mean = ln(0.84), SD = ln(0.86 / 0.84) / 1.96)
		80 - 84	log-normal (mean = ln(0.87), SD = ln(0.9 / 0.87) / 1.96)
BMI[74]	Both	30 - 59	log-normal (mean = ln(1.21), SD = ln(1.28 / 1.21) / 1.96)
		60 - 69	log-normal (mean = ln(1.06), SD = ln(1.12 / 1.06) / 1.96)
Diabetes Mellitus[284]	Both	40 - 59	log-normal (mean = ln(2.51), SD = ln(2.8/ 2.51) / 1.96)
		60 - 69	log-normal (mean = ln(2.01), SD = ln(2.26/ 2.01) / 1.96)
		70 - 84	log-normal (mean = ln(1.78), SD = ln(2.05/ 1.78) / 1.96)

...continued

Variable[source]	Sex	Age group	Distribution
Physical activity[69]	Both	30 - 69	No active days: log-normal (mean = ln(1.71), SD = ln(1.85/ 1.71) / 1.96)
			1 – 4 active days: log-normal (mean = ln(1.44), SD = ln(1.62/ 1.44) / 1.96)
		70 - 79	No active days: log-normal (mean = ln(1.5), SD = ln(1.61/ 1.5) / 1.96)
			1 – 4 active days: log-normal (mean = ln(1.31), SD = ln(1.48/ 1.31) / 1.96)
		80 - 84	No active days: log-normal (mean = ln(1.4), SD = ln(1.41/ 1.4) / 1.96)
			1 – 4 active days: log-normal (mean = ln(1.2), SD = ln(1.35/ 1.2) / 1.96)
Fruit and vegetables[285]	Both	30 - 84	log-normal (mean = ln(0.96), SD = ln(1.0.99/ 0.96) / 1.96)
Relative risks of relevant risk factors for stroke			
Active smoking[37]	Men	30 - 59	log-normal (mean = ln(3.12), SD = ln(4.64 / 3.12) / 1.96)
		60 - 69	log-normal (mean = ln(1.87), SD = ln(2.44 / 1.87) / 1.96)
		70 - 79	log-normal (mean = ln(1.39), SD = ln(1.77 / 1.39) / 1.96)
	Women	30 - 59	log-normal (mean = ln(4.61), SD = ln(6.37 / 4.61) / 1.96)

…continued

Variable[source]	Sex	Age group	Distribution
		60 - 69	log-normal (mean = ln(2.81), SD = ln(3.58 / 2.81) / 1.96)
		70 - 79	log-normal (mean = ln(1.95), SD = ln(2.45 / 1.95) / 1.96)
Environmental tobacco smoking[281]	Both	30 - 84	log-normal (mean = ln(1.25), SD = ln(1.38 / 1.25) / 1.96)
SBP[78]	Men	30 - 49	log-normal (mean = ln(0.33), SD = ln(0.38 / 0.33) / 1.96)
		50 - 59	log-normal (mean = ln(0.34), SD = ln(0.37 / 0.34) / 1.96)
		60 - 69	log-normal (mean = ln(0.41), SD = ln(0.44 / 0.41) / 1.96)
		70 - 74	log-normal (mean = ln(0.48), SD = ln(0.51 / 0.48) / 1.96)
		80 - 84	log-normal (mean = ln(0.68), SD = ln(0.75 / 0.68) / 1.96)
	Women	30 - 49	log-normal (mean = ln(0.41), SD = ln(0.49 / 0.41) / 1.96)
		50 - 59	log-normal (mean = ln(0.45), SD = ln(0.5 / 0.45) / 1.96)
		60 - 69	log-normal (mean = ln(0.47), SD = ln(0.51 / 0.47) / 1.96)

…continued

Variable[source]	Sex	Age group	Distribution
		70 - 74	log-normal (mean = ln(0.53), SD = ln(0.56 / 0.53) / 1.96)
		80 - 84	log-normal (mean = ln(0.65), SD = ln(0.71 / 0.65) / 1.96)
Total cholestero[81]	Both	40 - 49	log-normal (mean = ln(0.87), SD = ln(1 / 0.87) / 1.96)
		50 - 59	log-normal (mean = ln(0.91), SD = ln(0.97 / 0.91) / 1.96)
		60 - 69	log-normal (mean = ln(0.93), SD = ln(0.97 / 0.93) / 1.96)
BMI[74]	Both	30 - 59	log-normal (mean = ln(1.18), SD = ln(1.26 / 1.18) / 1.96)
		60 - 69	log-normal (mean = ln(1.08), SD = ln(1.15 / 1.08) / 1.96)
Diabetes mellitus[284]	Both	40 - 59	log-normal (mean = ln(3.74), SD = ln(4.58/ 3.74) / 1.96)
		60 - 69	log-normal (mean = ln(2.06), SD = ln(2.58/ 2.06) / 1.96)
		70 - 84	log-normal (mean = ln(1.8), SD = ln(2.27/ 1.8) / 1.96)
Physical activity[69]	Both	30 - 69	No active days: log-normal (mean = ln(1.53), SD = ln(1.79/ 1.53 / 1.96)

Variable[source]	Sex	Age group	Distribution
		70 - 79	No active days: log-normal (mean = ln(1.38), SD = ln(1.6/ 1.38) / 1.96)
		80 - 84	No active days: log-normal (mean = ln(1.24), SD = ln(1.45/ 1.24) / 1.96)
Fruit and vegetables[286]	Both	30 - 84	log-normal (mean = ln(0.95), SD = ln(0.97/ 0.95) / 1.96)
Relative risks of relevant risk factors for lung cancer			
Environmental tobacco smoking[282]	Both	30 - 84	log-normal (mean = ln(1.33), SD = ln(1.54/ 1.33) / 1.96)
Fruit and vegetables[52]	Both	30 - 84	log-normal (mean = ln(0.96^0.8), SD =ln(0.98^0.8/0.96^0.8) / 1.96)
Relative risks of relevant risk factors for gstric cancer			
Active smoking (duration in years)[278]	Both	30 - 84	Normal (mean = 0.03, SD = 0.002)
Ex-smoking (years since cessation)[278]	Both	30 - 84	log-normal (mean = ln(0.96), SD = ln(1/ 0.96) / 1.96)
BMI[24]	Both	30 - 84	Normal (mean and SD is a function of BMI)
Fruit and vegetables[48]	Both	30 - 69	log-normal (mean = ln(0.94), SD = ln(1/ 0.94) / 1.96)
	Both	70 - 79	log-normal (mean = ln(0.96), SD = ln(1/ 0.96) / 1.96)
	Both	80 - 84	log-normal (mean = ln(0.97), SD = ln(1/ 0.97) / 1.96)

...continued

Variable[source]	Sex	Age group	Distribution
Salt[31]	Both	30 - 84	log-normal (mean = ln(1.08), SD = ln(1.08/ 1) / 1.96)
Universal screening scenario and variations (chapter 5 on page 121)			
Proportion of participants with a QRISK2 score between 10% and 20%[370]	Both	40 - 74	PERT (min = 0.2, mode = 0.25, max = 0.3, shape = 4)
Proportion of participants with a QRISK2 score higher than 20%[370]	Both	40 - 74	PERT (min = 0.04, mode = 0.05, max = 0.1, shape = 4)
Atorvastatin 20mg relative reduction on total cholesterol[323, 325]	Both	40 - 74	Normal (mean = 0.32, sd = 0.14)
Atorvastatin prescription uptake[469] (QRISK2: 10% - 20%)	Both	40 - 74	PERT (min = 0.07, mode = 0.17, max = 0.24, shape = 4)
Atorvastatin prescription uptake[469] (QRISK2: >20%)	Both	40 - 74	PERT (min = 0.2, mode = 0.24, max = 0.3, shape = 4)
Antihypertensive medication prescription uptake[469] (QRISK2: 10% - 20%)	Both	40 - 74	PERT (min = 0.05, mode = 0.13, max = 0.2, shape = 4)
Antihypertensive medication prescription uptake[469] (QRISK2: >20%)	Both	40 - 74	PERT (min = 0.15, mode = 0.23, max = 0.3, shape = 4)
Persistence with medication[352]	Both	40 - 74	PERT (min = 0.5, mode = 0.8, max = 1, shape = 4)
Adherence to medication[352]	Both	40 - 74	PERT (min = 0.3, mode = 0.7, max = 1, shape = 4)
First year smoking cessation success rate[378, 379]	Both	40 - 74	Bernoulli (probability = 0.1)

Variable[source]	Sex	Age group	Distribution
Relative BMI reduction	Both	40 - 74	1 - PERT (min = 0.97, mode = 0.99, max = 1, shape = 4)
Proportion of high-risk participants to increase F&V consumption by a portion per day	Both	40 - 74	Bernoulli (probability = 0.5)
Proportion of high-risk participants to increase PA by a day per week	Both	40 - 74	Bernoulli (probability = 0.5)
Universal screening scenario (20% treatment threshold) (chapter 5 on page 121)			
Atorvastatin prescription uptake[469] (QRISK2: 10% - 20%)	Both	40 - 74	PERT (min = 0.01, mode = 0.07, max = 0.10, shape = 4)
Population-wide intervention (chapter 5 on page 121)			
Smoking prevalence relative reduction[173]	Both	30 - 84	PERT (min = 0.05, mode = 0.13, max = 0.14, shape = 4)
BMI rate of increase relative reduction[380–382]	Both	30 - 84	1 - PERT (min = 0.98, mode = 0.99, max = 1, shape = 4)
SBP absolute decrease (mmHg)[176, 470]	Both	30 - 84	PERT (min = 0.18, mode = 0.81, max = 1.10, shape = 4)
Proportion of the population to increase their F&V consumption by one portion[384, 385]	Both	30 - 84	PERT (min = 0.2, mode = 0.5, max = 0.8, shape = 4)
Other inputs			
CVD lag time	Both	30 - 84	1 + Binomial(n = 9, p = (5-1)/9)
Cancer lag time	Both	30 - 84	1 + Binomial(n = 9, p = (8-1)/9)

...continued

Variable[source]	Sex	Age group	Distribution
Optimal salt consumption[59]	Both	30 - 84	PERT(min = 1.5, mode = 3.8, max = 6, shape = 4)
Stricter salt policy target	Both	30 - 84	PERT(min = 5.8, mode = 6, max = 7, shape = 4)

Abbreviations: body mass index (BMI), cardiovascular disease (CVD), coronary heart disease (CHD), fruit and vegetable (F&V), physical activity (PA), standard deviation (SD), systolic blood pressure (SBP).

B

VALIDATION APPENDIX

B.1 SYNTHETIC POPULATION VALIDATION
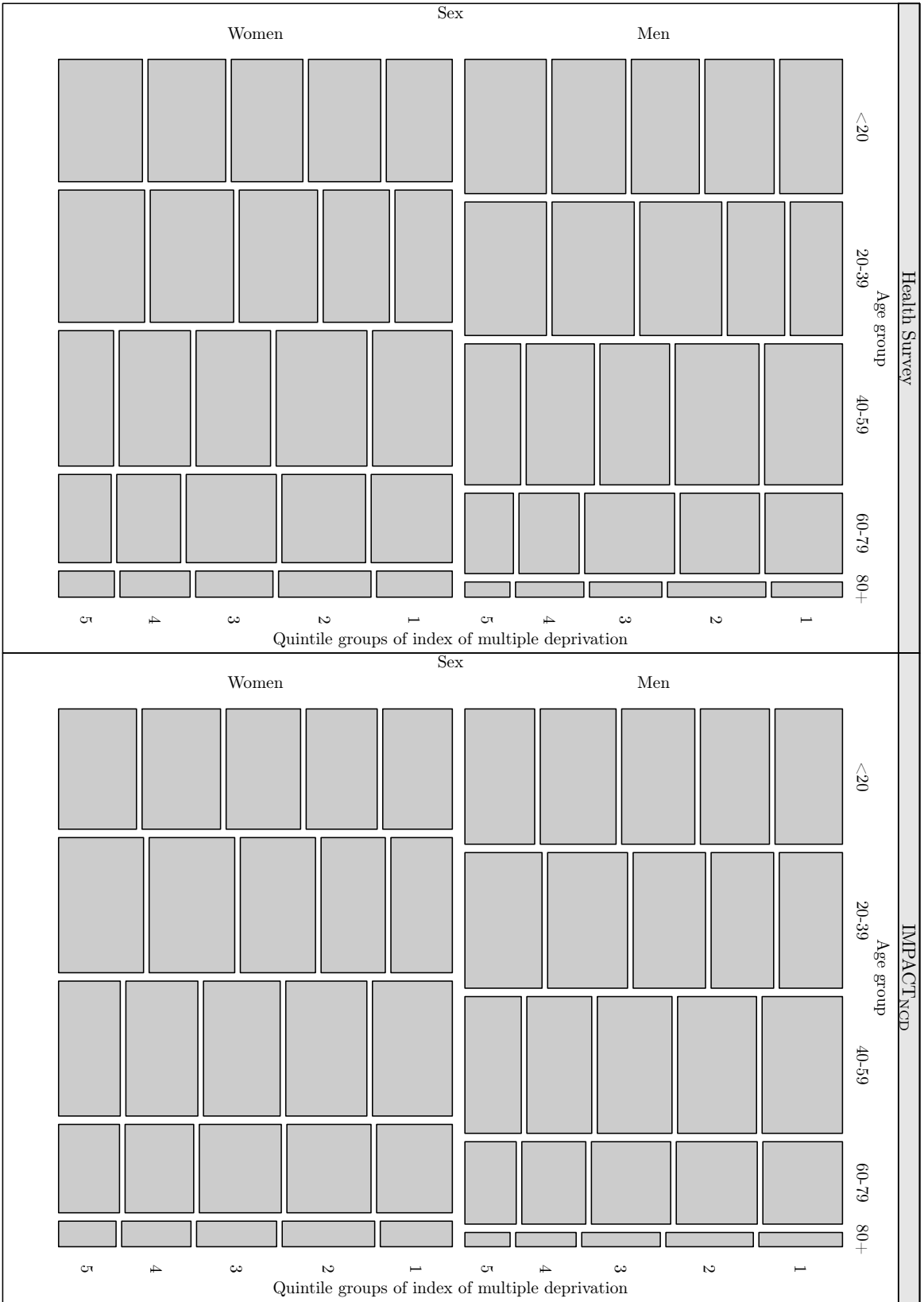
Figure B.1: Comparison between the Health Survey for England 2006 and a random sample from the synthetic population. Distribution of age group, sex and quintile groups of Index of Multiple Deprivation (1 = least deprived, 5 = most deprived) is presented.
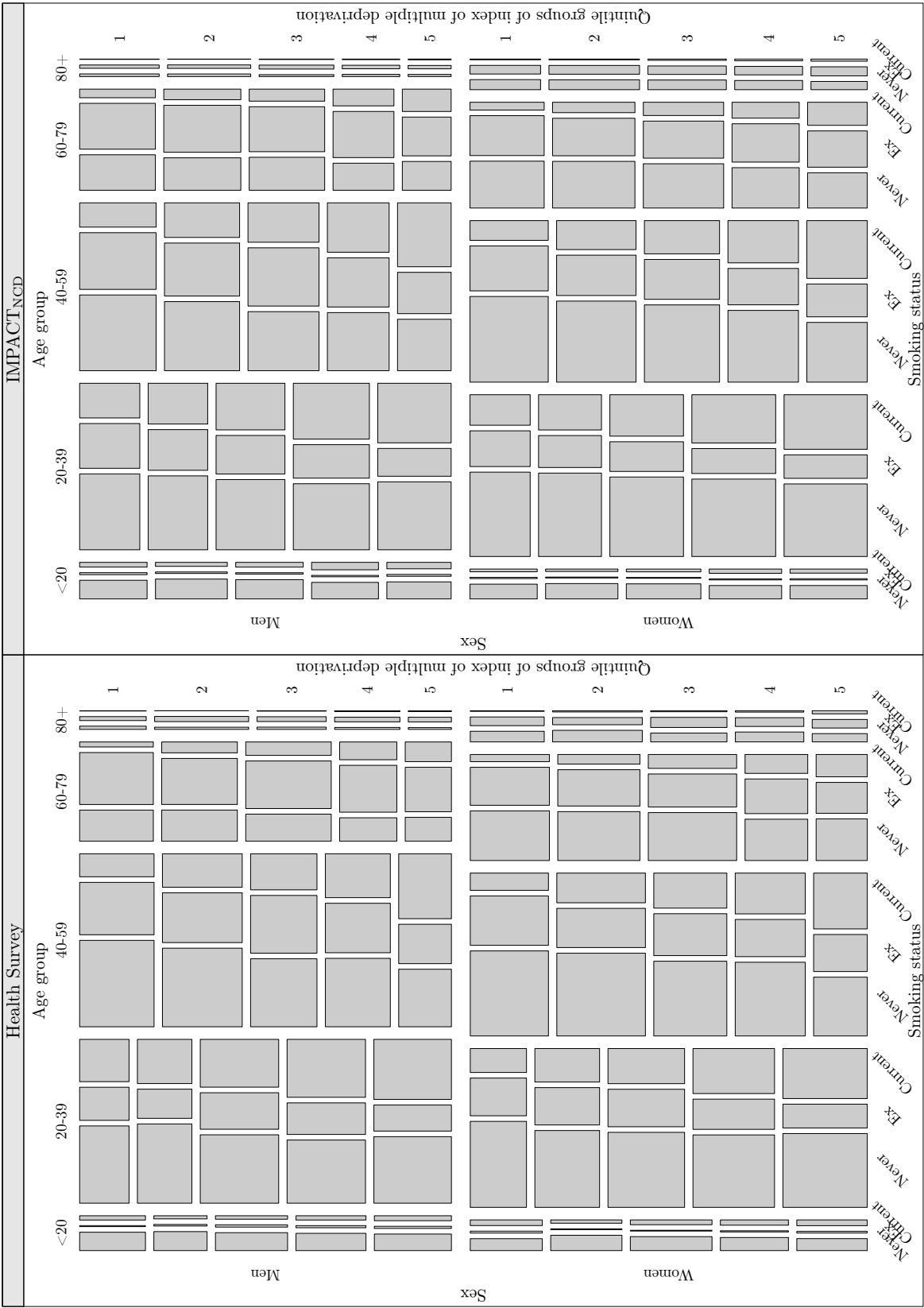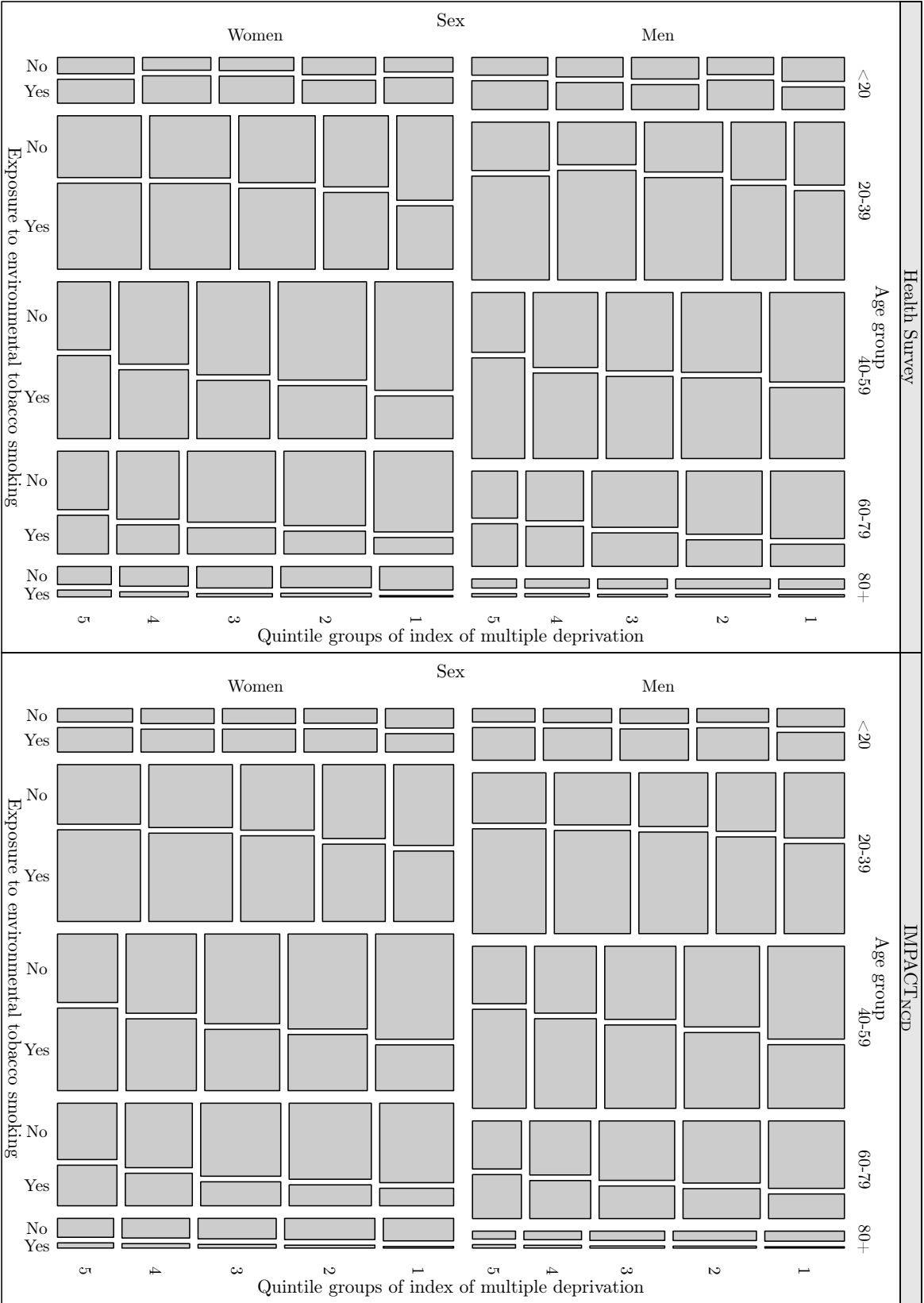
Figure B.2: Comparison between the Health Survey for England 2006 and a random sample from the synthetic population. Distribution of age group, sex, quintile groups of Index of Multiple Deprivation (1 = least deprived, 5 = most deprived) and smoking status is presented.
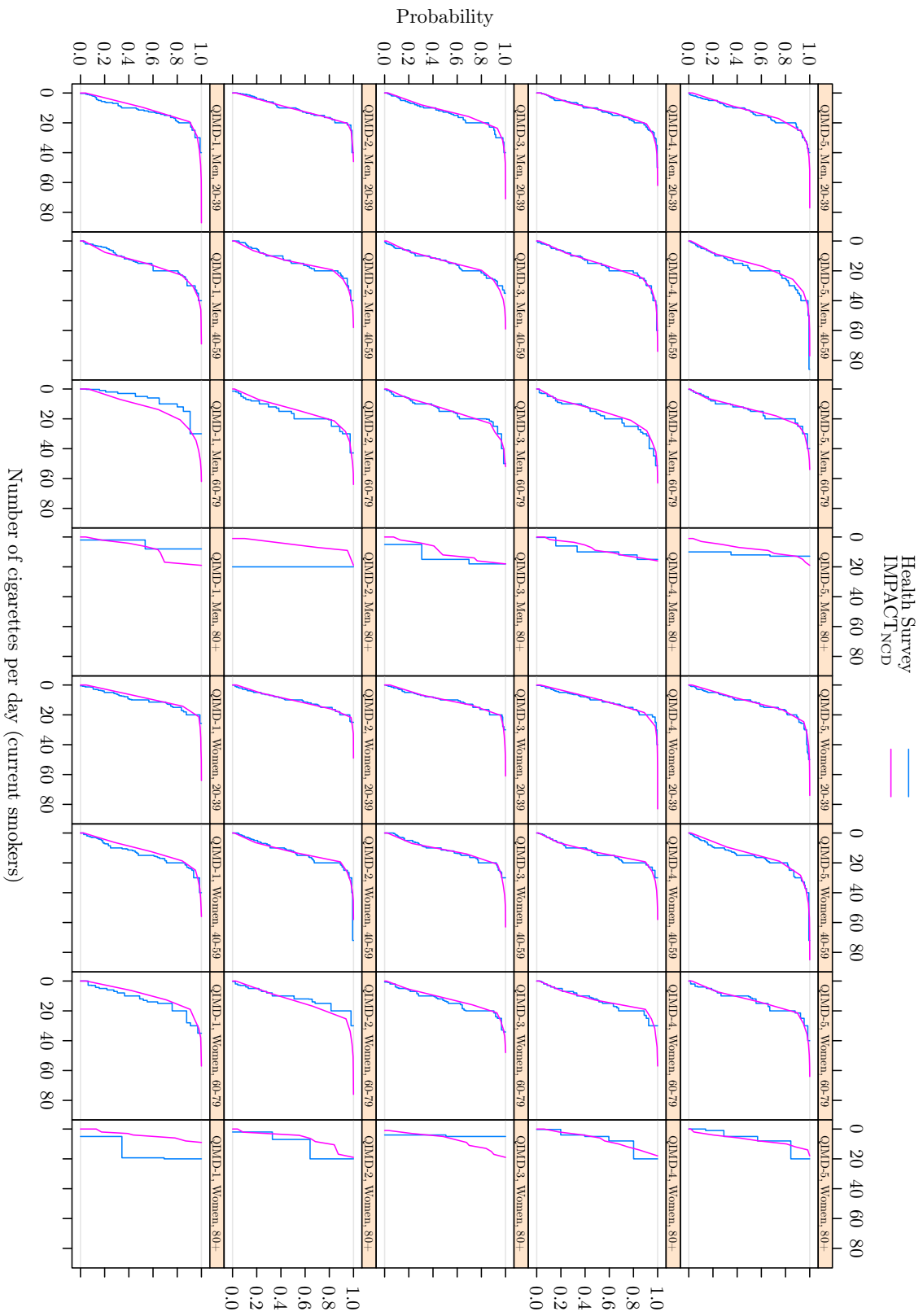
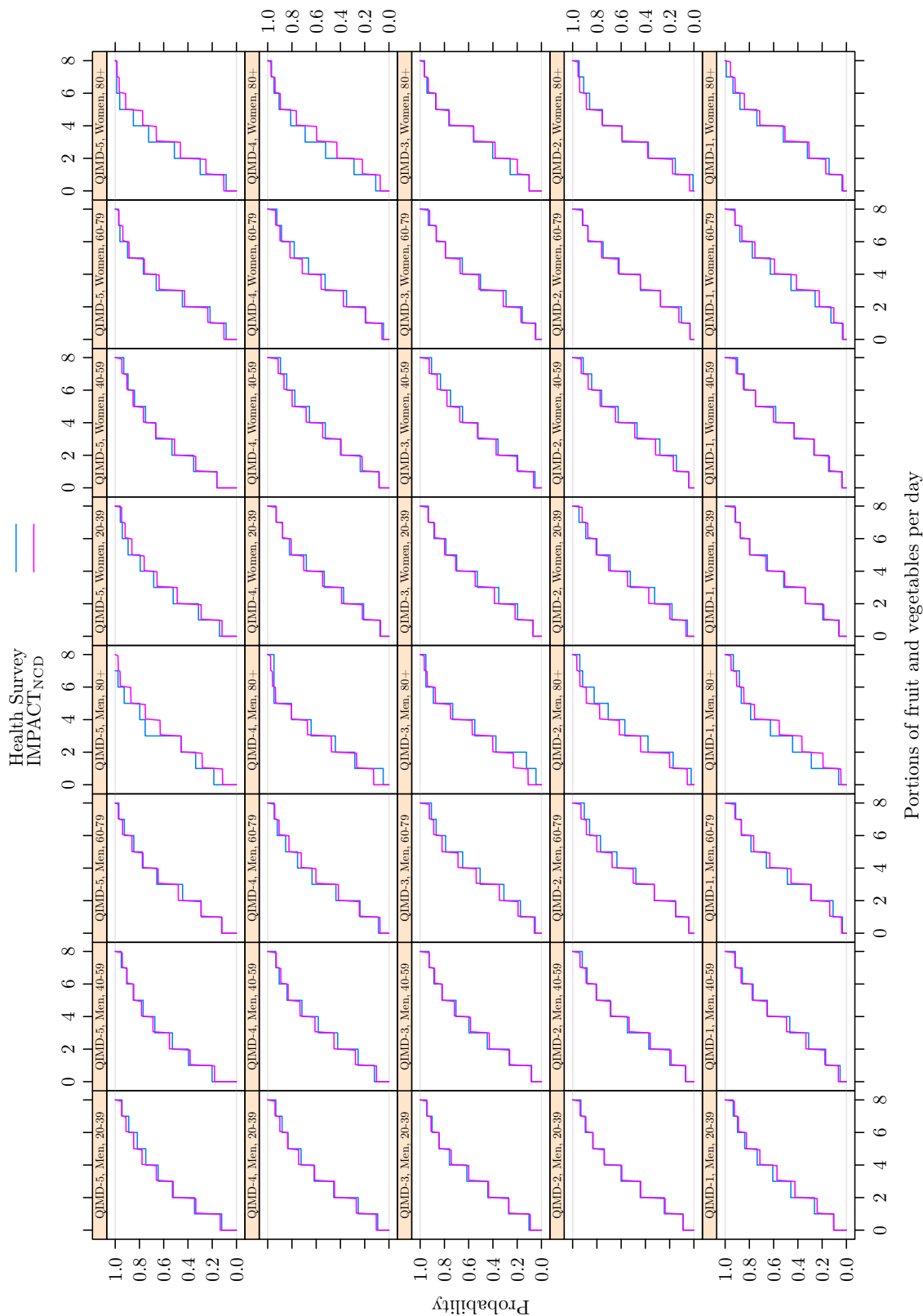Figure B.3: Comparison of smoking intensity cumulative distributions for current smokers in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.
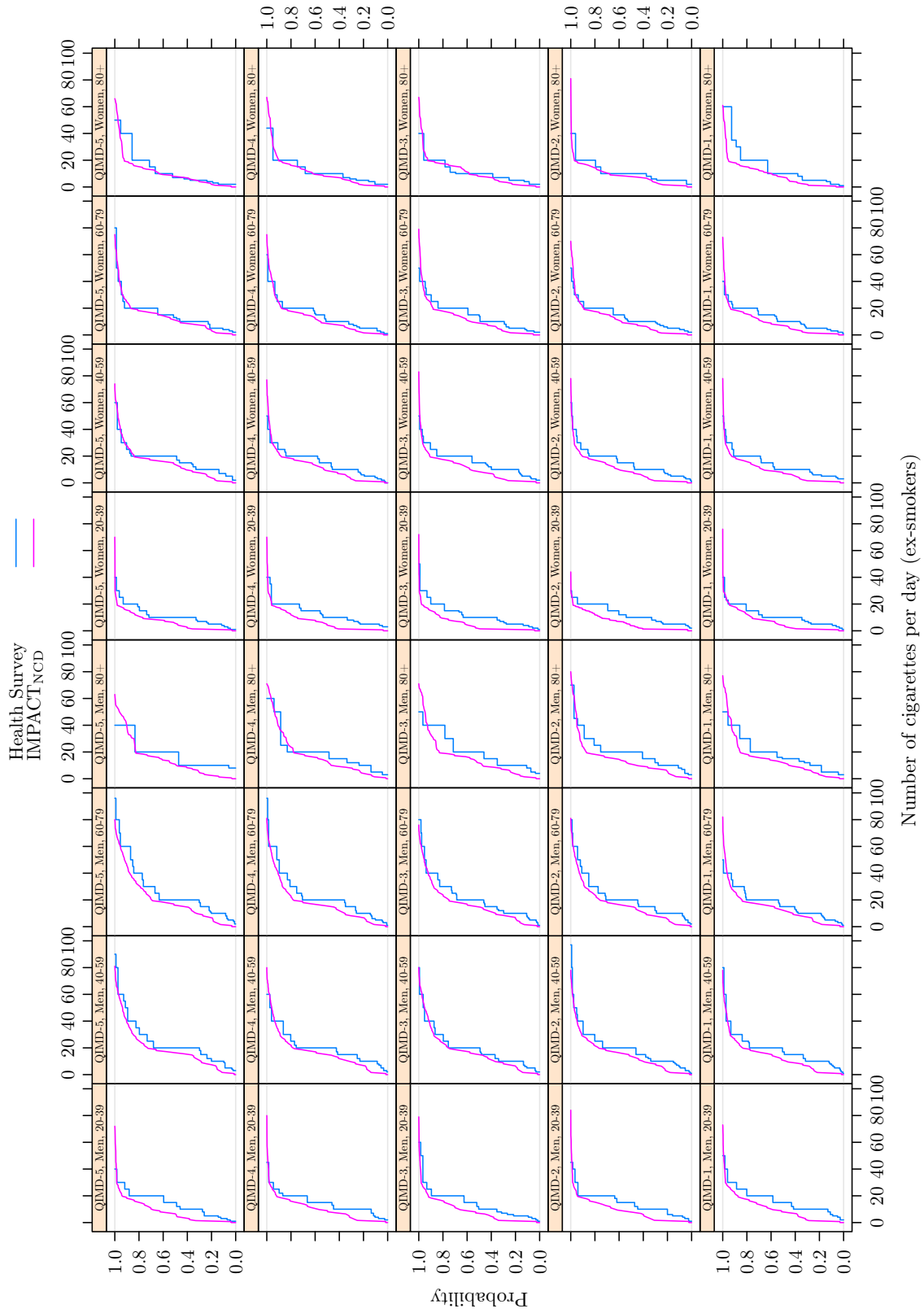
Figure B.4: Comparison of smoking intensity cumulative distributions for ex-smokers in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.
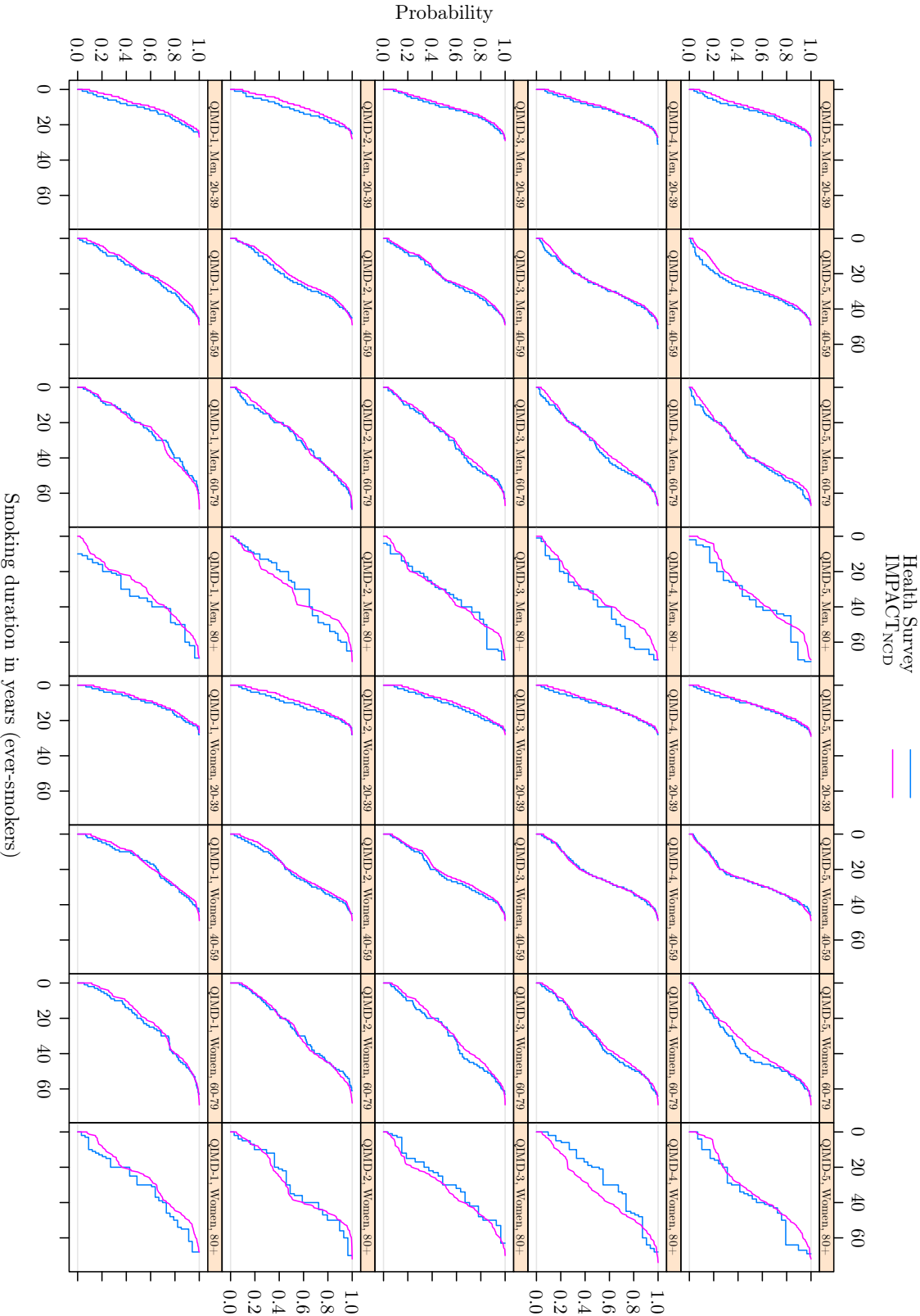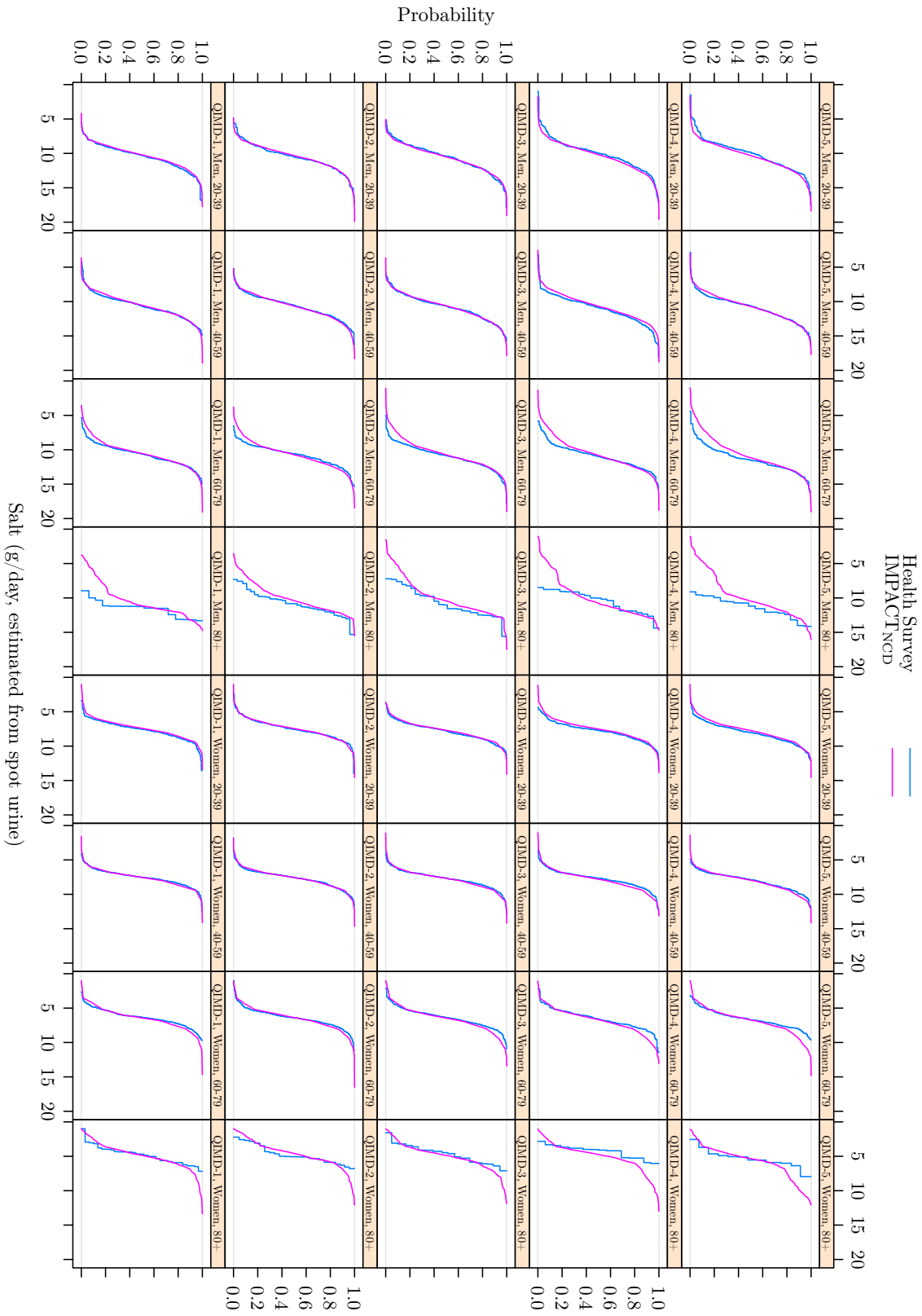
Figure B.5: Comparison of smoking duration cumulative distributions for ever-smokers in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.
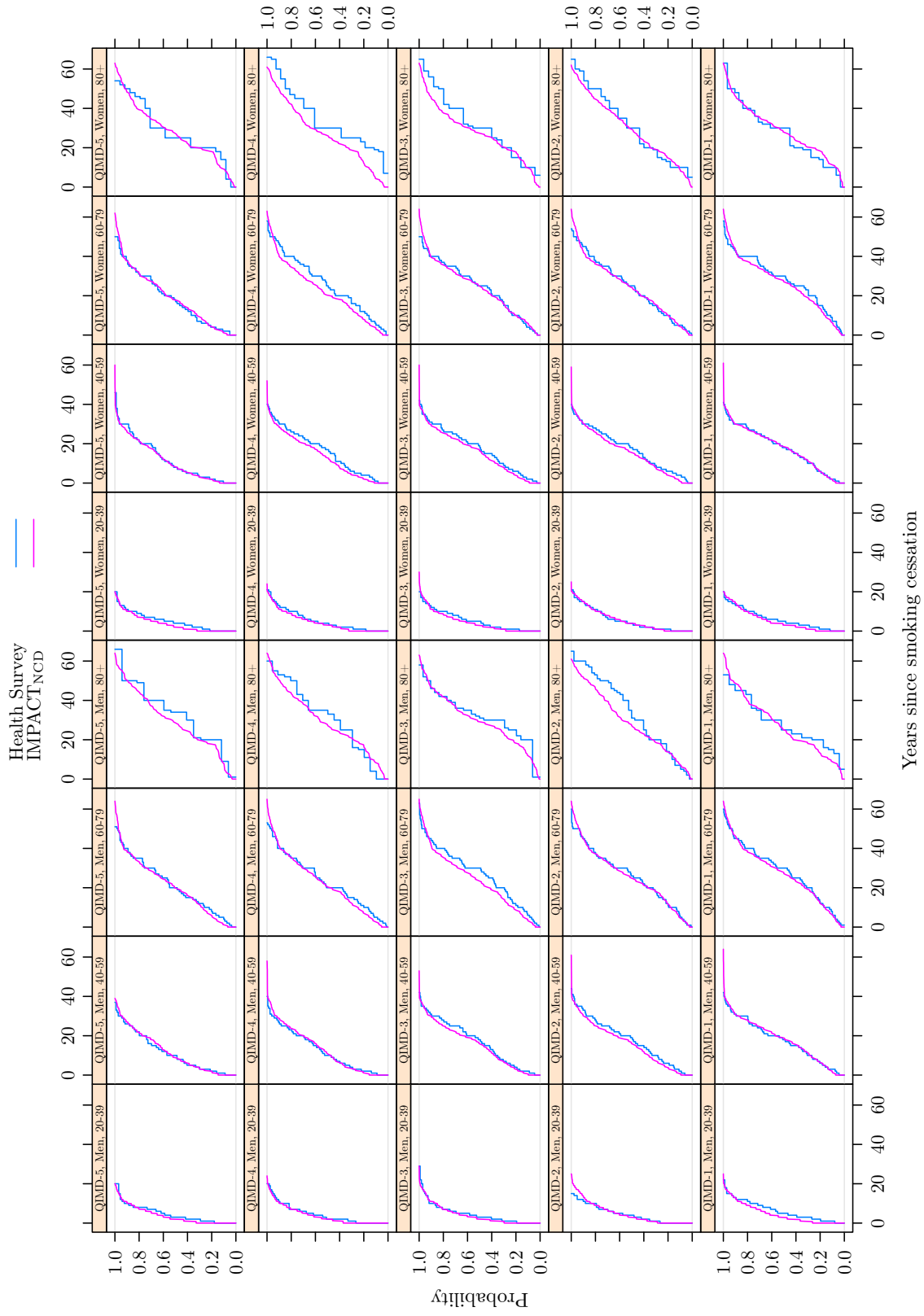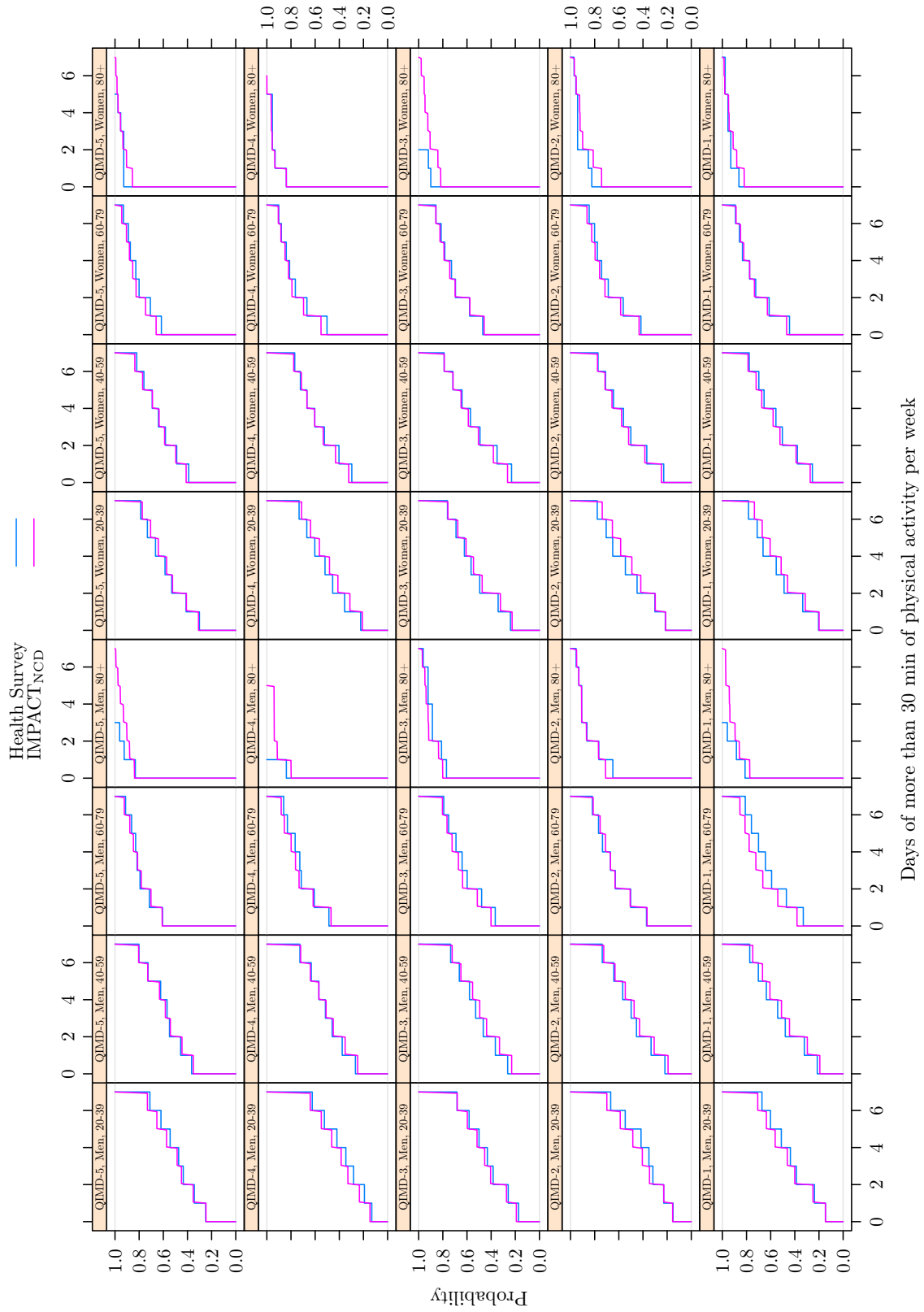
Figure B.6: Comparison of years since smoking cessation cumulative distributions for ex-smokers in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.
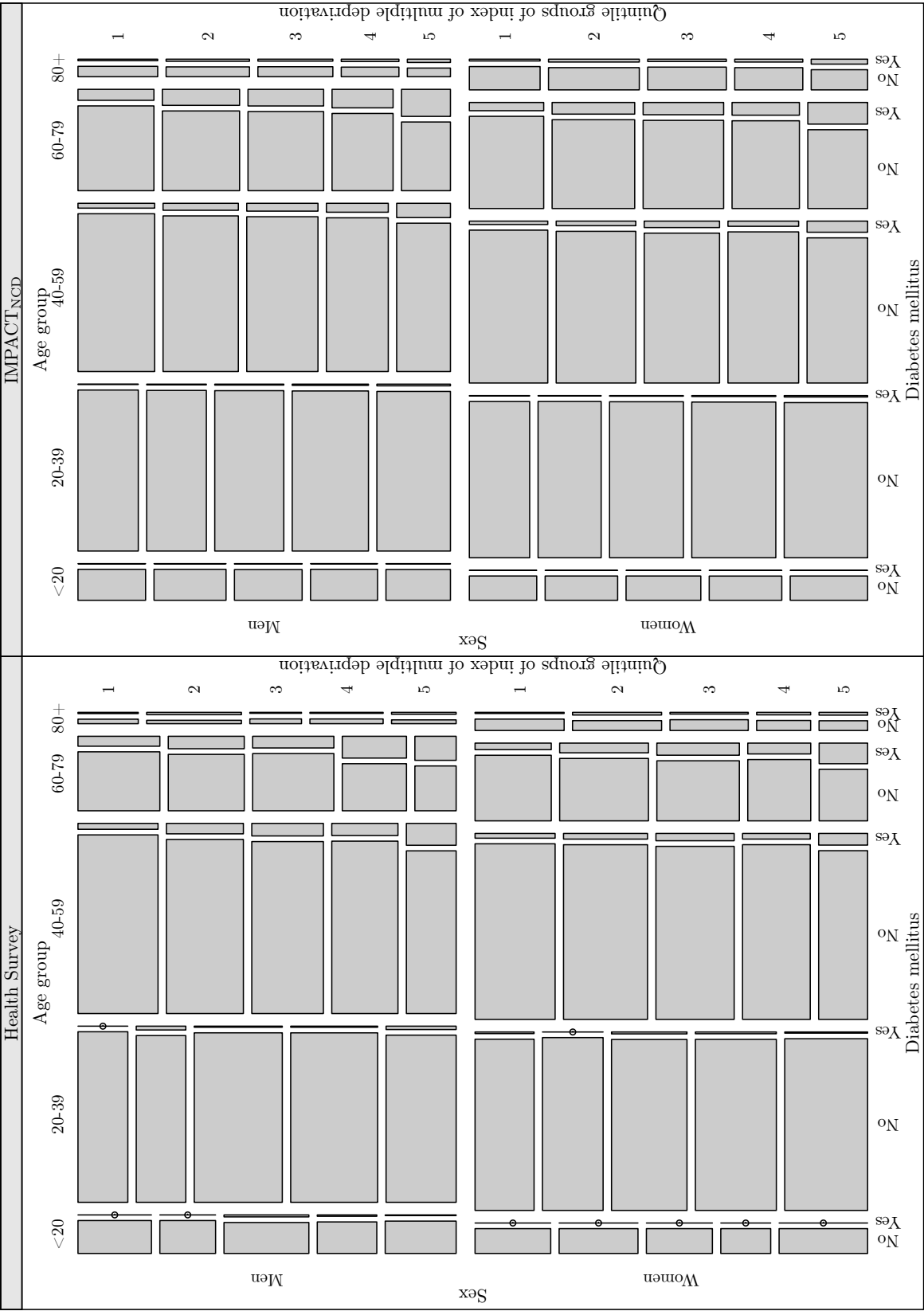
Figure B.7: Comparison between the Health Survey for England 2006 and a random sample from the synthetic population. Distribution of age group, sex, quintile groups of Index of Multiple Deprivation (1 = least deprived, 5 = most deprived) and exposure to environmental tobacco is presented.

Figure B.8: Comparison of fruit and vegetable consumption cumulative distributions in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex, and age group.

Figure B.9: Comparison of salt consumption (from spot urine) cumulative distributions in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.
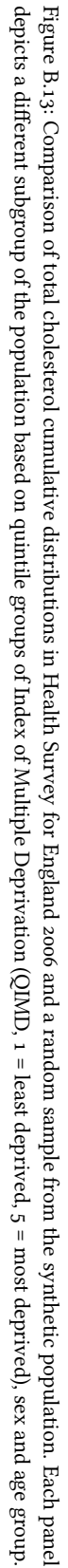
Figure B.10: Comparison of physical activity cumulative distributions in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex, and age group.
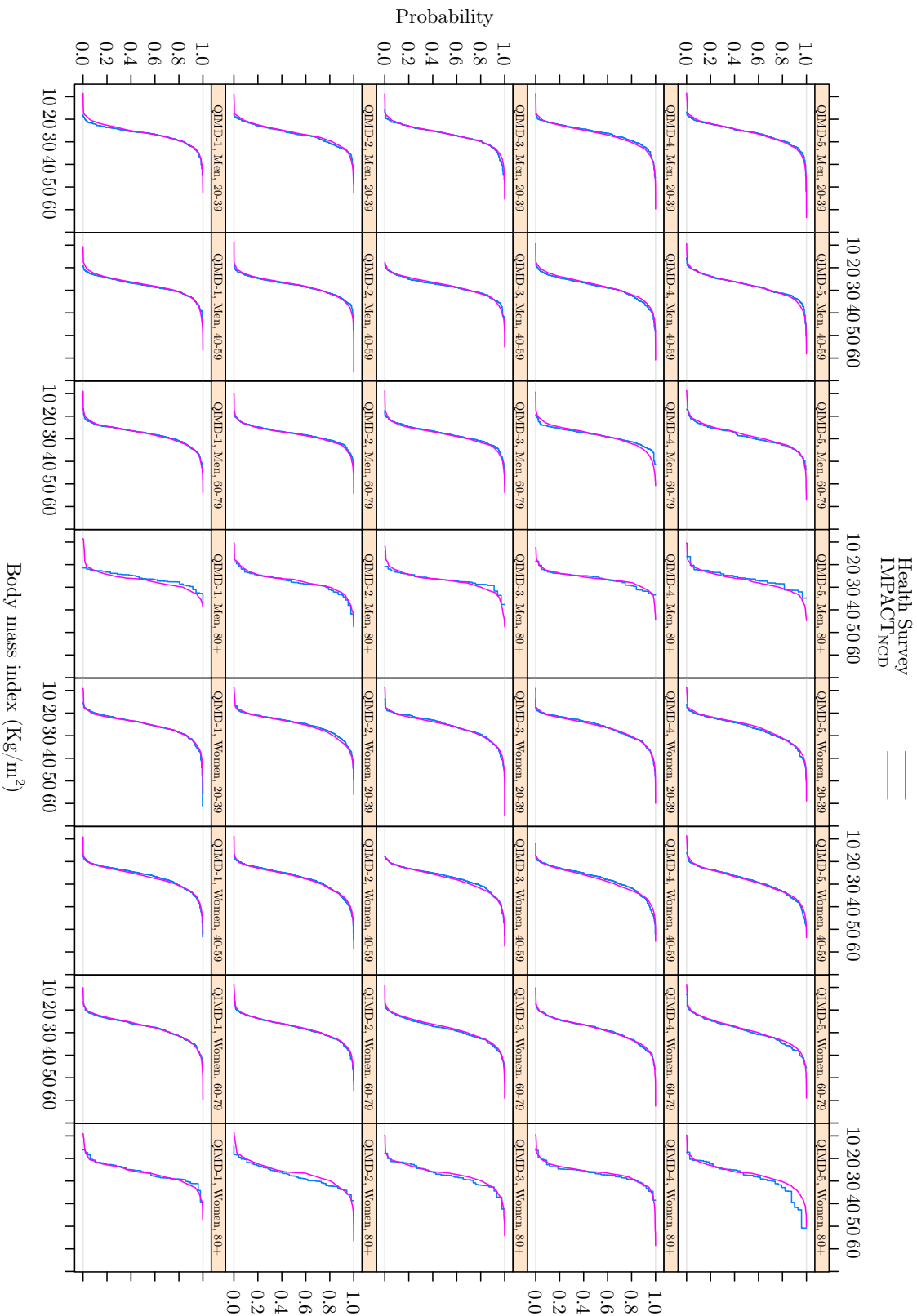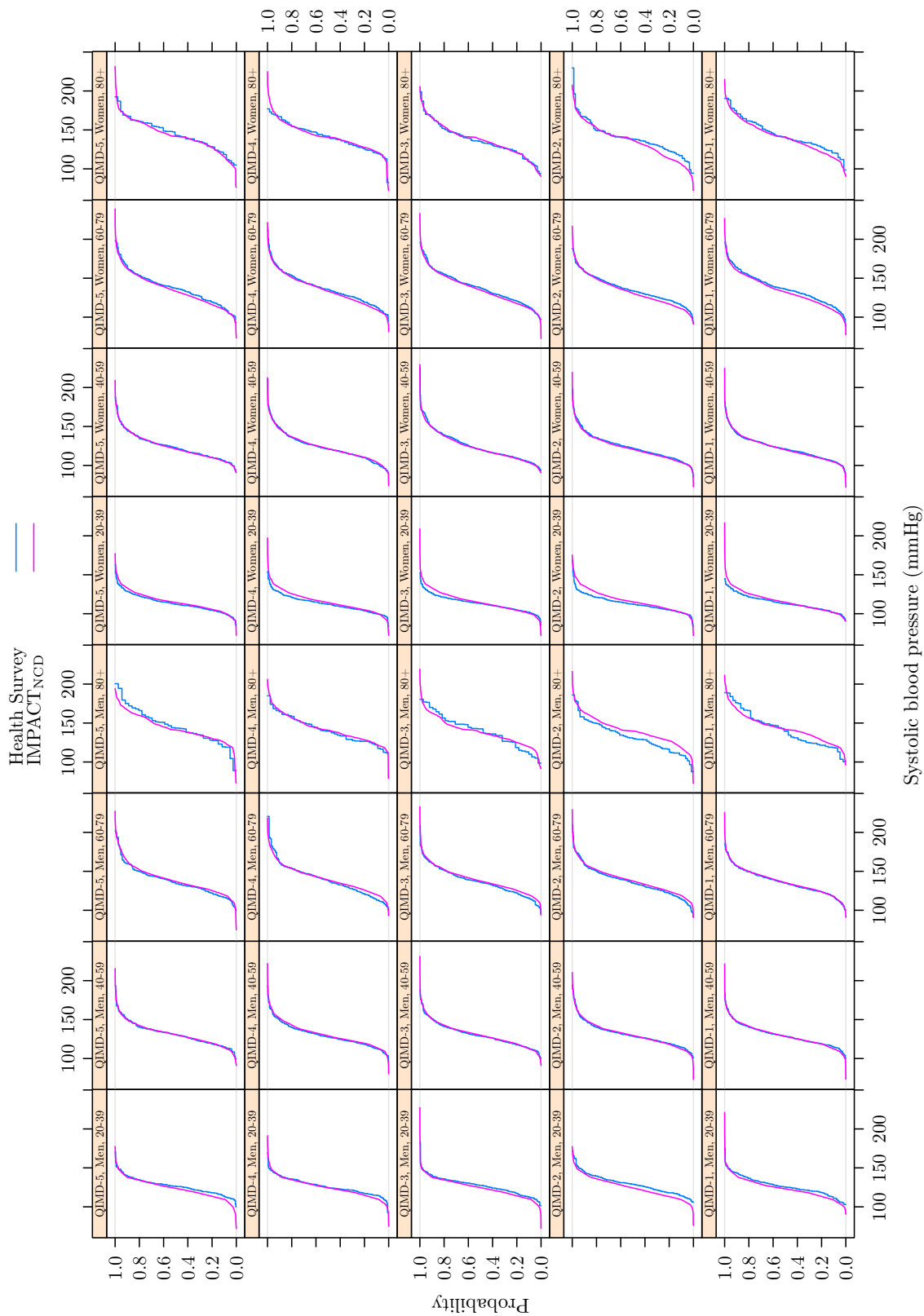
Figure B.11: Comparison of body mass index cumulative distributions in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.

Figure B.12: Comparison between the Health Survey for England 2006 and a random sample from the synthetic population. Distribution of age group, sex, quintile groups of Index of Multiple Deprivation (1 = least deprived, 5 = most deprived) and diabetes mellitus is presented. The small dots in the survey younger groups represent the lack of diabetic participants. The method I followed for population synthesis was able to produce diabetics in these groups.

Figure B.13: Comparison of total cholesterol cumulative distributions in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.

Figure B.14: Comparison of systolic blood pressure cumulative distributions in Health Survey for England 2006 and a random sample from the synthetic population. Each panel depicts a different subgroup of the population based on quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived), sex and age group.

	Age	Sex	QIMD	Smoking	ETS	F&V	PA	Diabetes	BMI	Cholesterol
Sex	.06									
QIMD	-.12	-.01								
Smoking	.22	-.10	.09							
ETS	-.36	-.18	.19	.10						
F&V	.09	.08	-.15	-.02	-.17					
PA	-.26	-.15	-.04	-.01	.14	.14				
Diabetes	.49	-.12	.10	.12	-.15	.03	-.33			
BMI	.27	-.05	.02	.08	-.04	.02	-.15	.30		
Cholesterol	.47	.04	-.07	.11	-.13	.02	-.10	.03	.26	
SBP	.42	-.23	-.02	.09	-.07	.01	-.09	.26	.29	.28

(a) Health Survey for England correlation structure.

	Age	Sex	QIMD	Smoking	ETS	F&V	PA	Diabetes	BMI	Cholesterol
Sex	.06									
QIMD	-.09	.03								
Smoking	.23	-.11	.09							
ETS	-.37	-.17	.18	.07						
F&V	.09	.07	-.14	.02	-.06					
PA	-.33	-.14	-.05	-.06	.12	-.00				
Diabetes	.49	-.08	.12	.12	-.14	.01	-.20			
BMI	.26	-.04	.03	.07	-.07	.03	-.04	.27		
Cholesterol	.32	.10	-.05	.07	-.12	.05	-.06	.11	.18	
SBP	.41	-.22	-.02	.09	-.11	.02	-.13	.24	.27	.12

(b) IMPACT$_{NCD}$ correlation structure.

	Age	Sex	QIMD	Smoking	ETS	F&V	PA	Diabetes	BMI	Cholesterol
Sex	.00									
QIMD	.03	.04								
Smoking	.01	-.01	-.00							
ETS	-.01	.01	-.01	-.02						
F&V	-.00	-.00	.01	.04	.11					
PA	-.07	.01	-.00	-.04	-.02	-.14				
Diabetes	-.00	.04	.02	.00	.00	-.01	.13			
BMI	-.01	.01	.00	-.01	-.03	.01	.11	-.03		
Cholesterol	-.15	.06	.02	-.03	.01	.03	.05	.08	-.08	
SBP	-.01	.01	.01	.01	-.05	.01	-.04	-.01	-.02	-.16

(c) Difference in the correlation structures (IMPACT$_{NCD}$- Health Survey for England).

Figure B.15: Comparison of correlations structures in Health Survey for England and a random sample from the synthetic population. Abbreviations: body mass index (BMI); environmental tobacco smoking (ETS); fruit and vegetable (F&V); quintile groups of Index of Multiple Deprivation (QIMD); systolic blood pressure (SBP).

B.2 RISK FACTOR TRENDS

HSE2005 appears to be an outlier in some graphs that risk factor exposures depend on blood tests. This is because in that survey year, bloods were only checked for those 65 and older and therefore are not representative of the population. In addition, it appears that IMPACT$_{NCD}$ underestimates the prevalence of smoking and diabetes mellitus. This is mostly an artefact. For computational efficiency IMPACT$_{NCD}$ only calculates risk factor exposures for the given lag times. For example, for a 50 year old synthetic individual in 2011 and with 5-year lag time IMPACT$_{NCD}$ estimates the smoking status for 2006 (2011 - 5) and for age of 45 (50 - 5). Therefore, IMPACT$_{NCD}$ smoking prevalence for 2006 is representative only for the part of the population that remained alive for five years, until 2011. This leads in survival bias because those less exposed to risk factors have a higher probability of remaining alive for the next five years; hence the reported IMPACT$_{NCD}$ risk exposures are underestimated. Because smokers and diabetics are modelled to have higher overall mortality (section 2.4.3 on page 59) the bias is more obvious in these two risk factor graphs. I would like to emphasise that the bias does not affect the outputs of IMPACT$_{NCD}$ and is only present in the validation graphs. Figure B.45 on page 243 was plotted assuming no lag time for smoking and it is apparent that survival bias was eliminated. As a side note, the effect of smoke free legislation is apparent in the environmental tobacco smoking graphs. I decided to model environmental tobacco smoking linearly and ignore this effect for simplicity. The introduced bias is small because of the small relative risk of environmental tobacco smoking that is usually around 1.2.
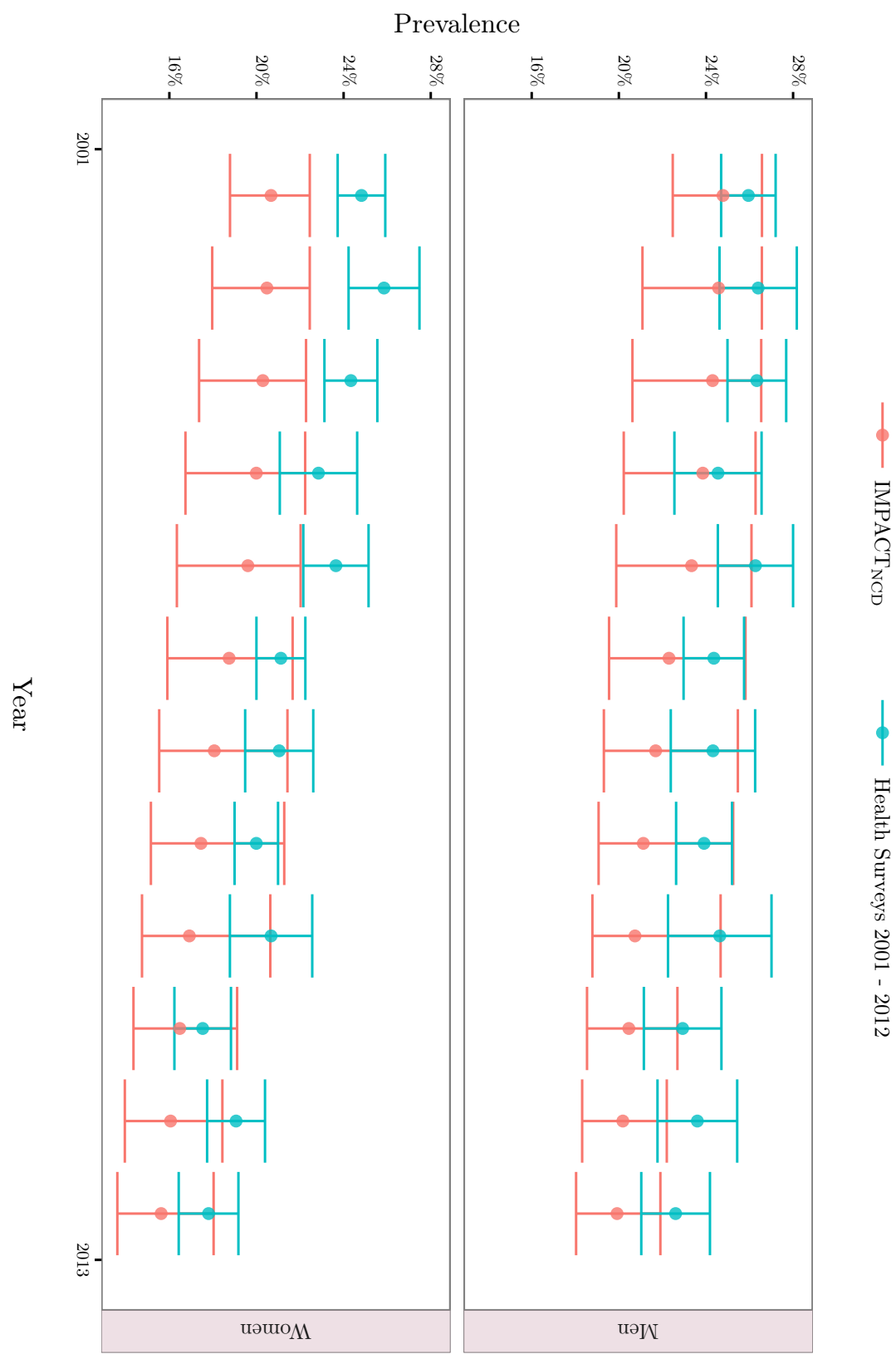
Figure B.16: Smoking prevalence for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT_NCD synthetic population estimates, stratified by sex. Error bars depict 95 % confidence intervals of the mean.
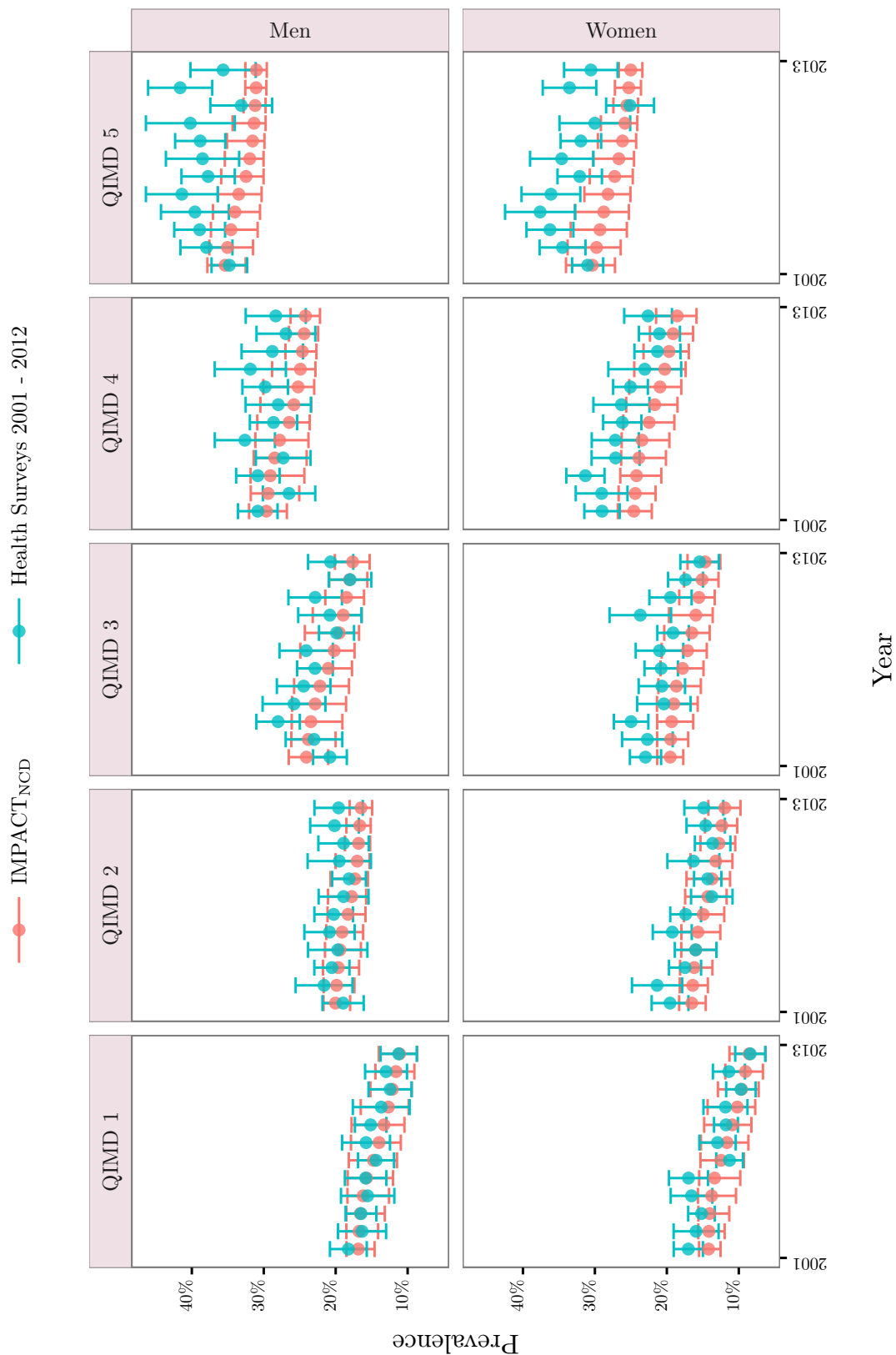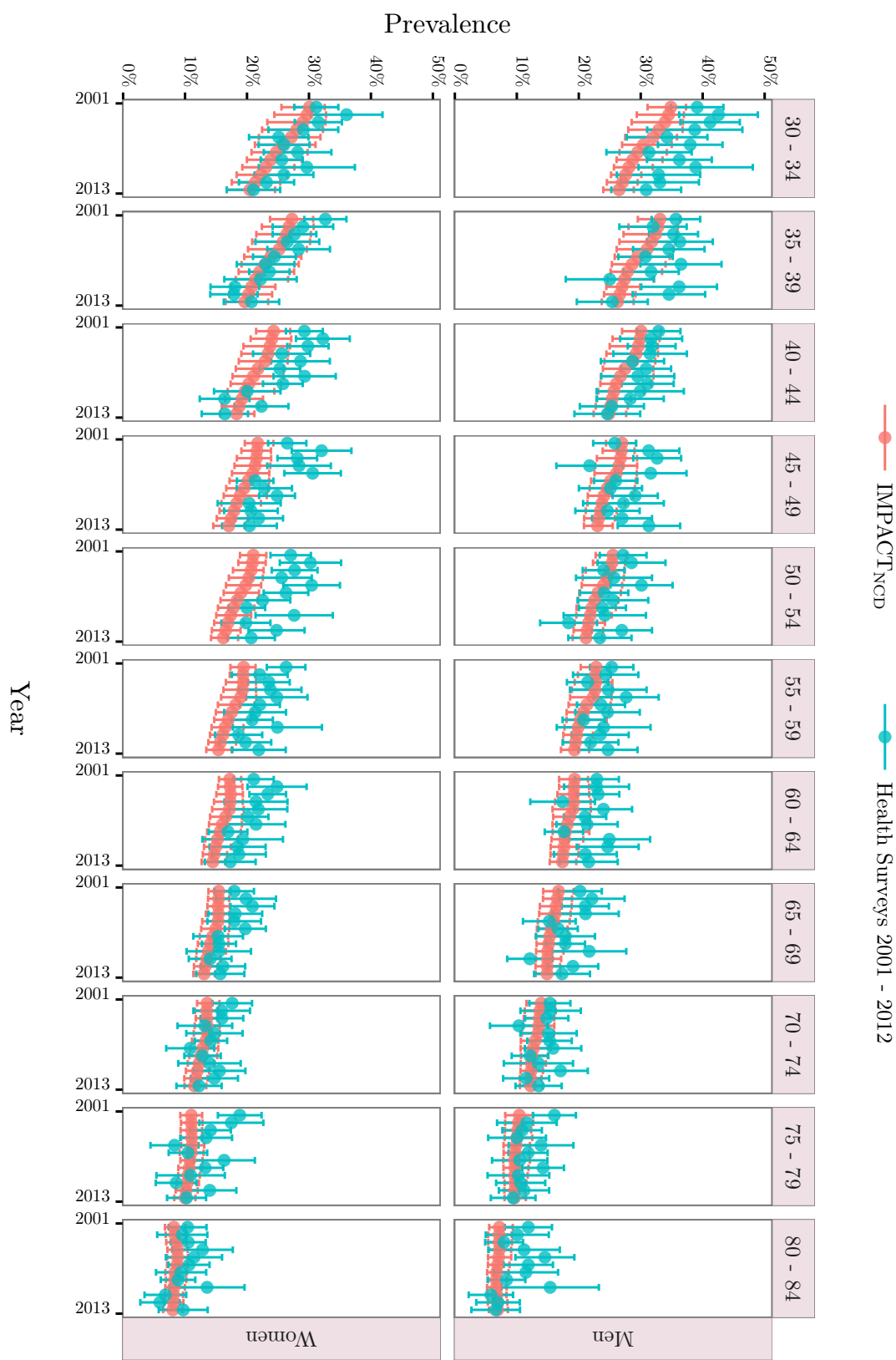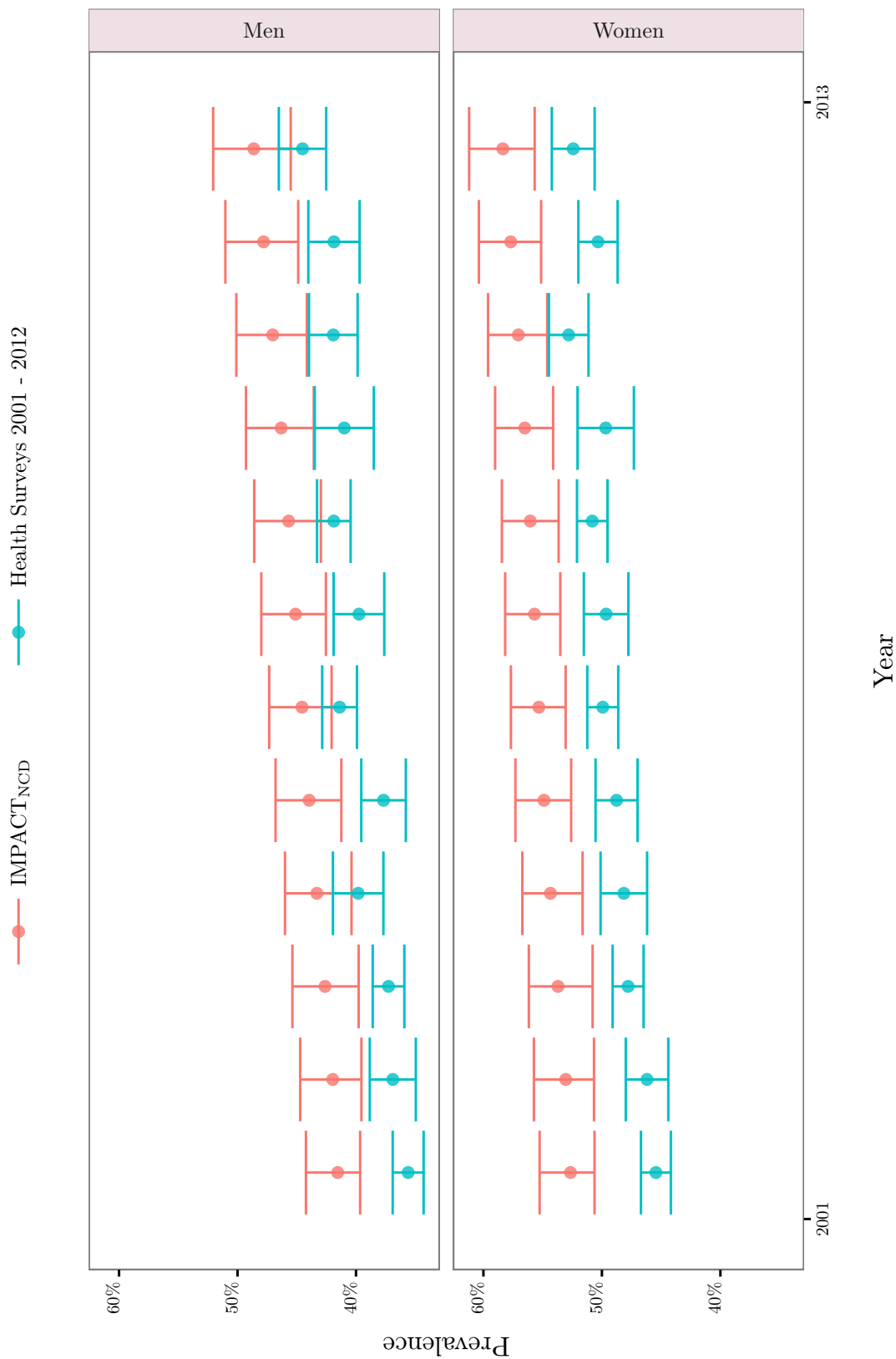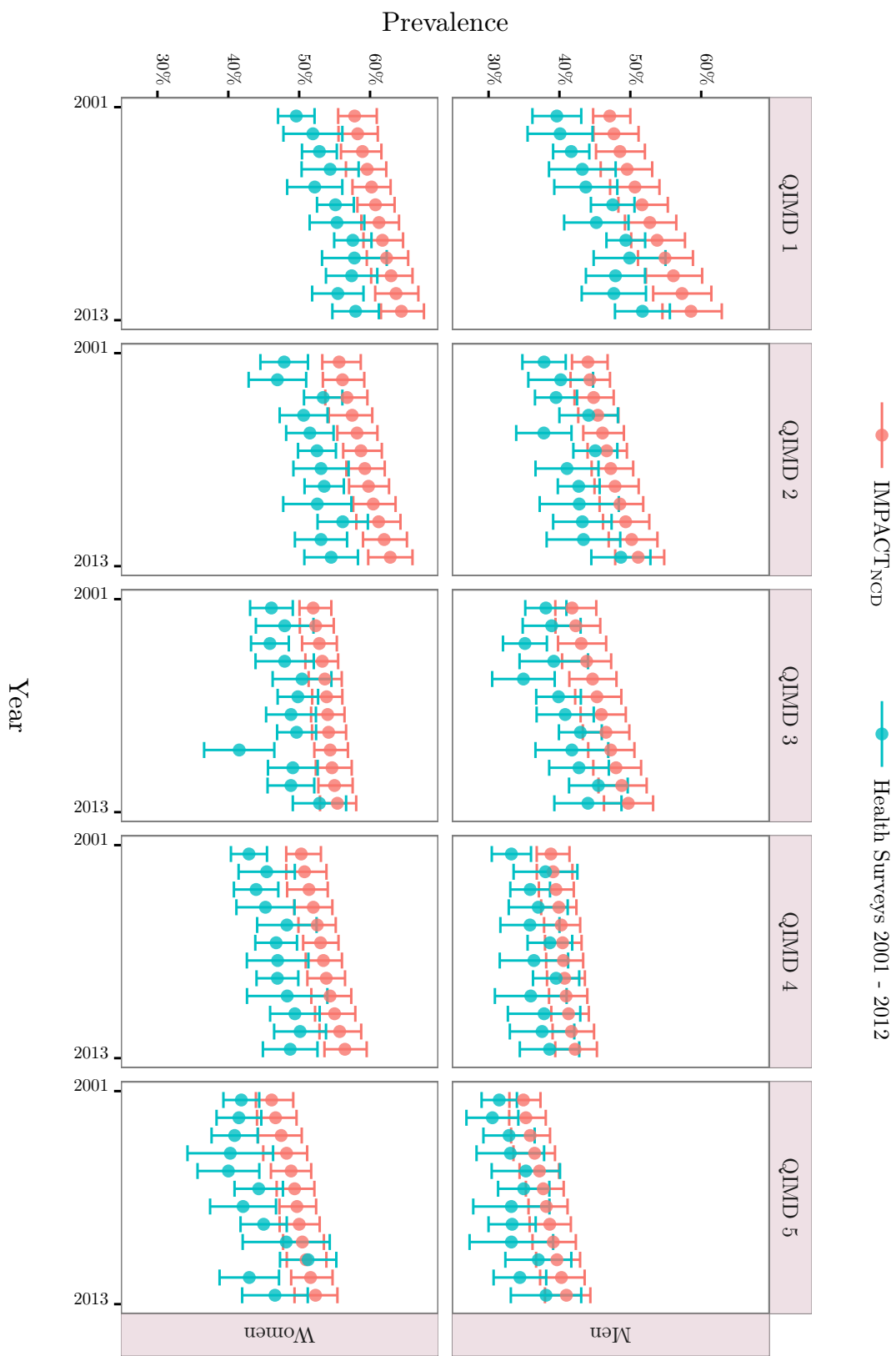
Figure B.17: Smoking prevalence for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
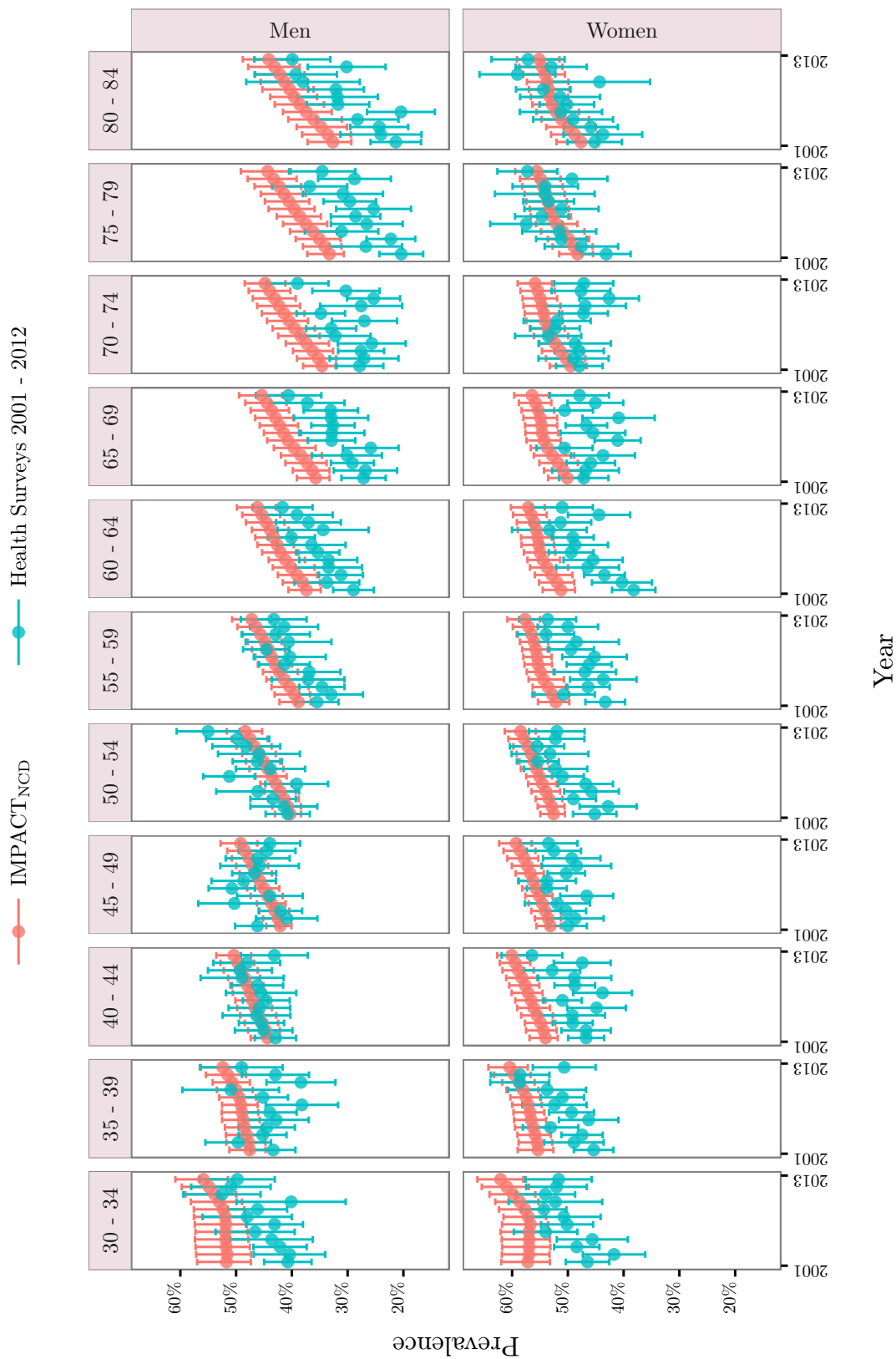
Figure B.18: Smoking prevalence for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.
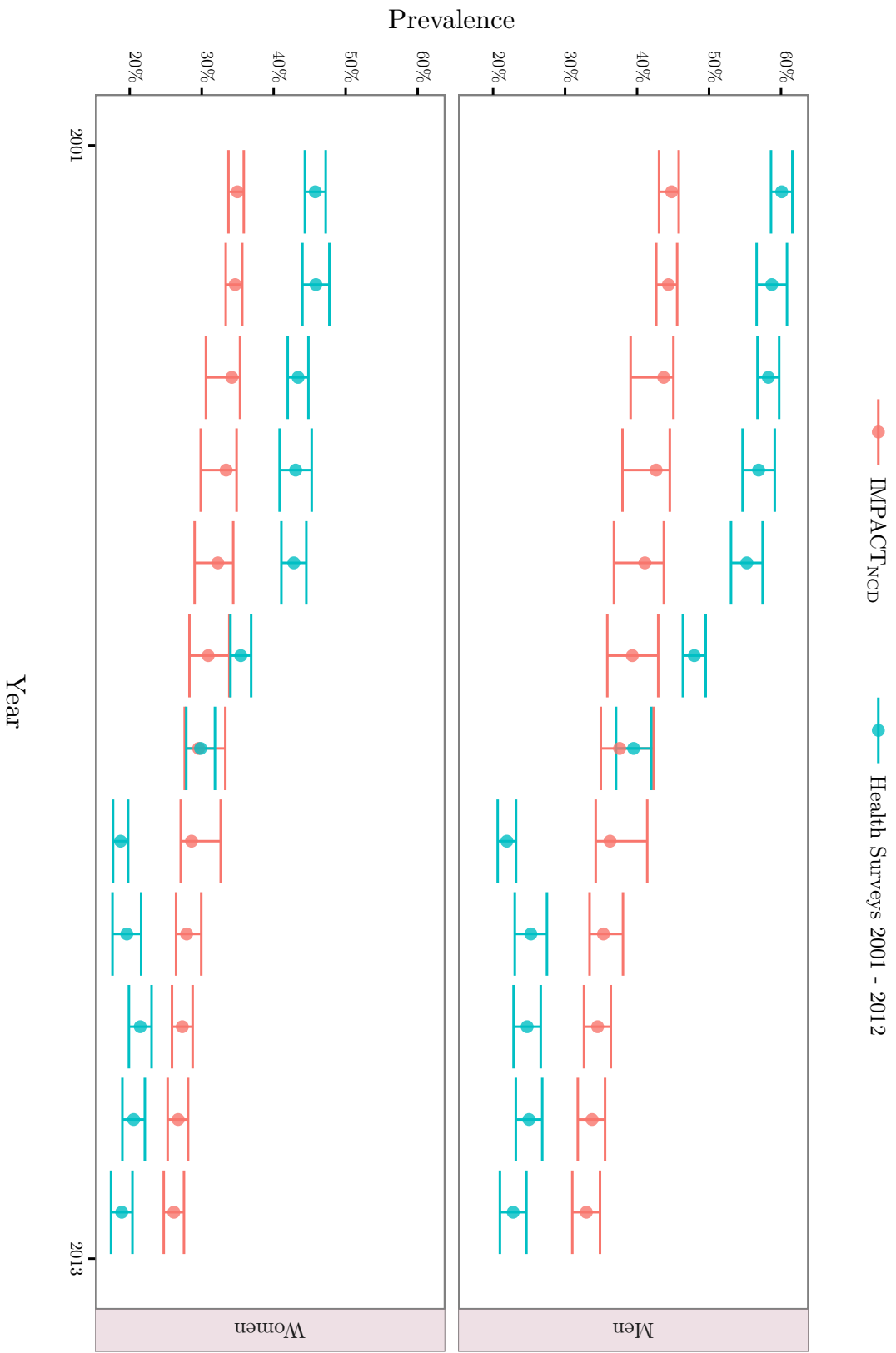
Figure B.19: Never-smoking for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.20: Never-smoking prevalence for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.21: Never-smoking prevalence for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.
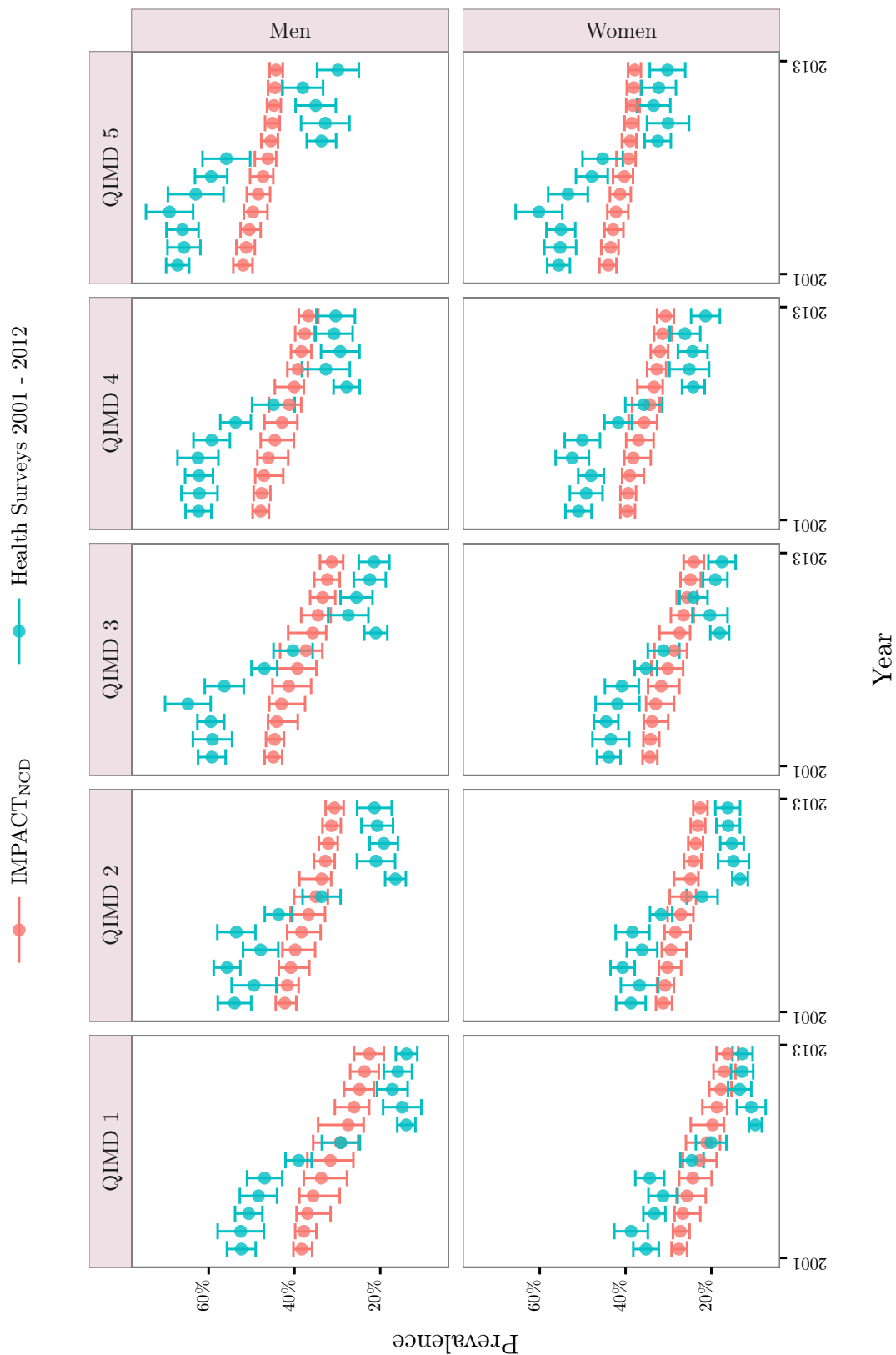
Figure B.22: Environmental tobacco smoking exposure prevalence for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.23: Environmental tobacco smoking exposure prevalence for ages 30 − 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
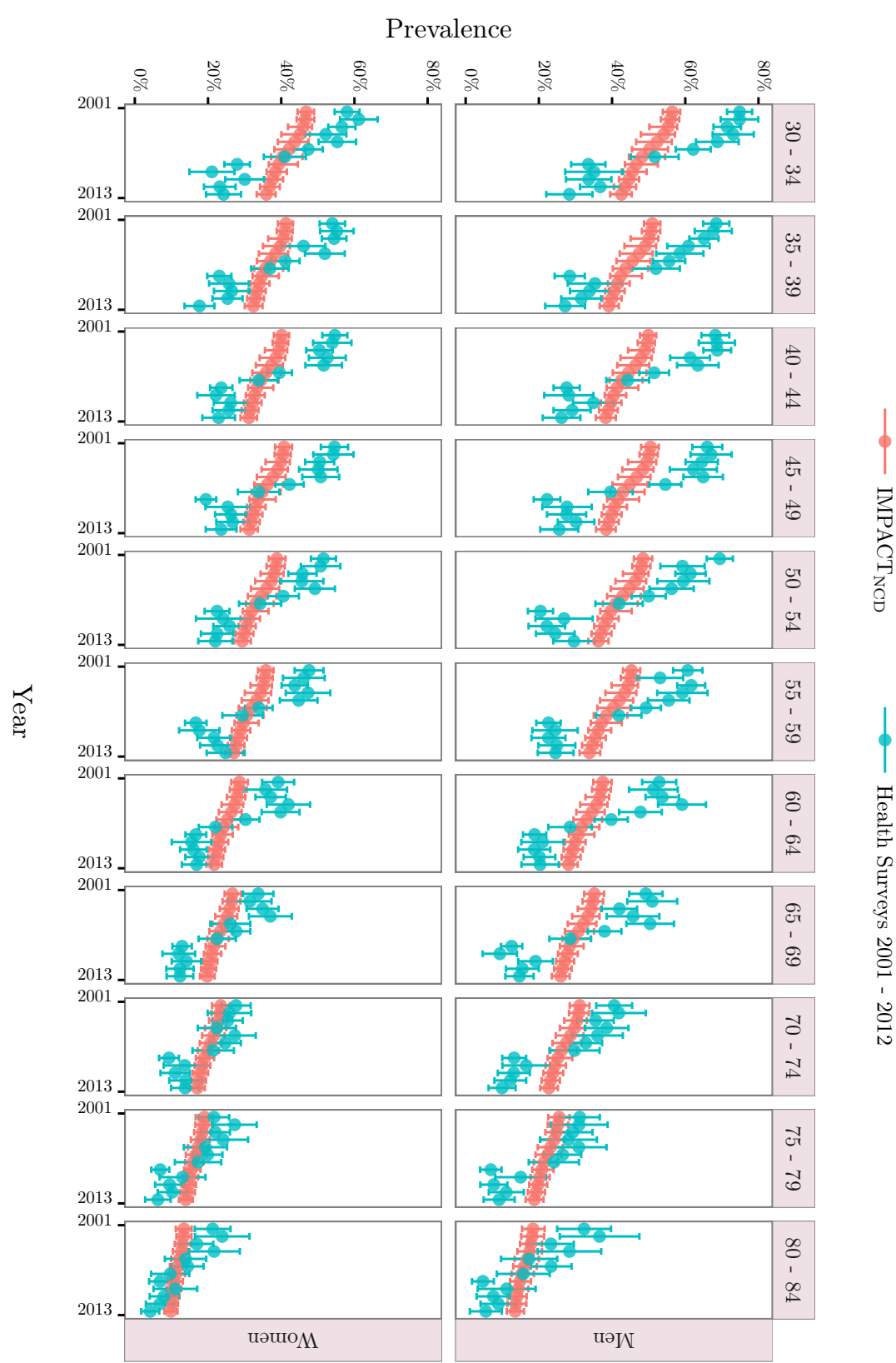
Figure B.24: Environmental tobacco smoking exposure prevalence for ages 30 − 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.
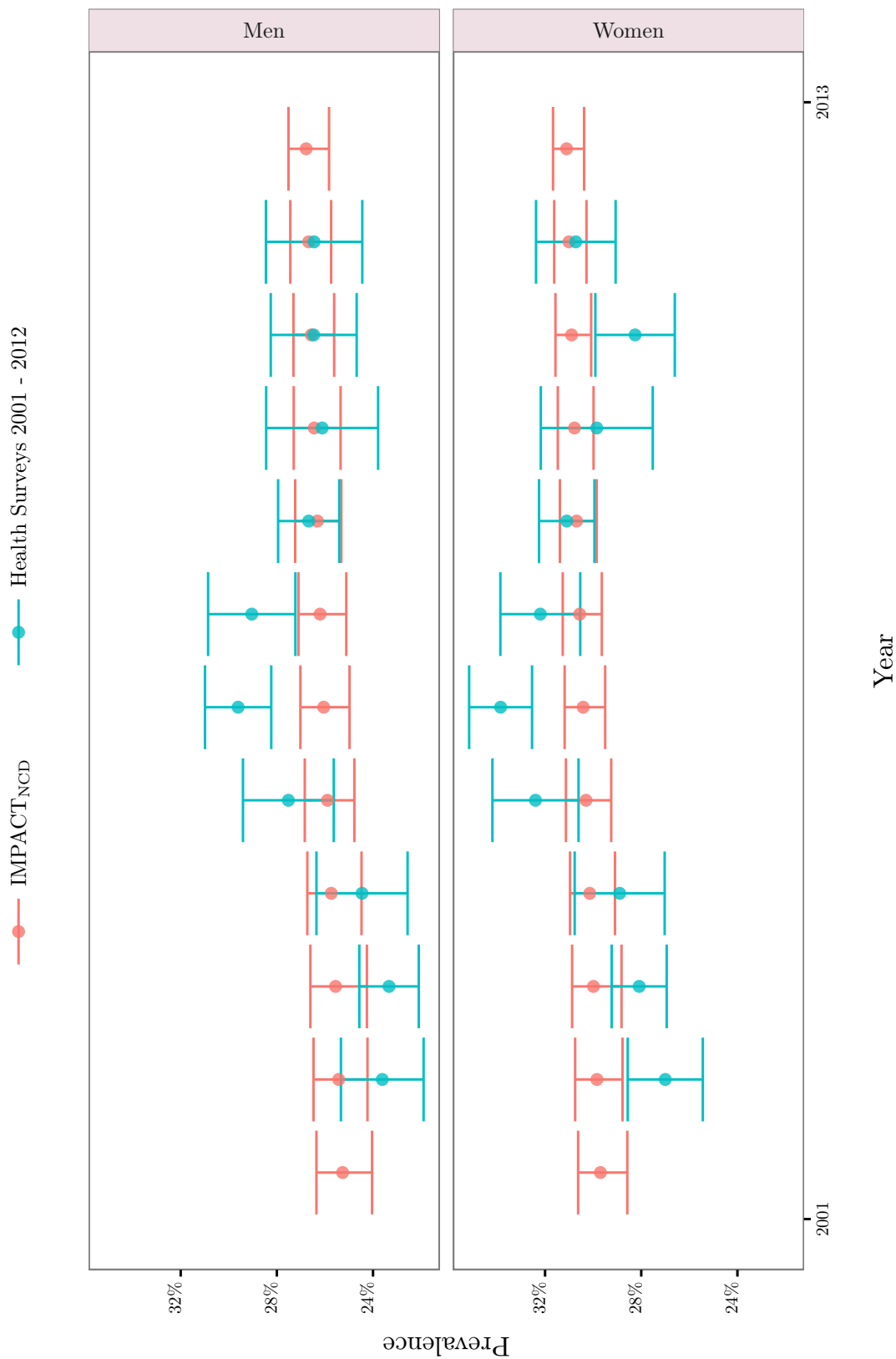
Figure B.25: Prevalence of consumption of five or more fruit and vegetable portions per day for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % confidence intervals of the mean.
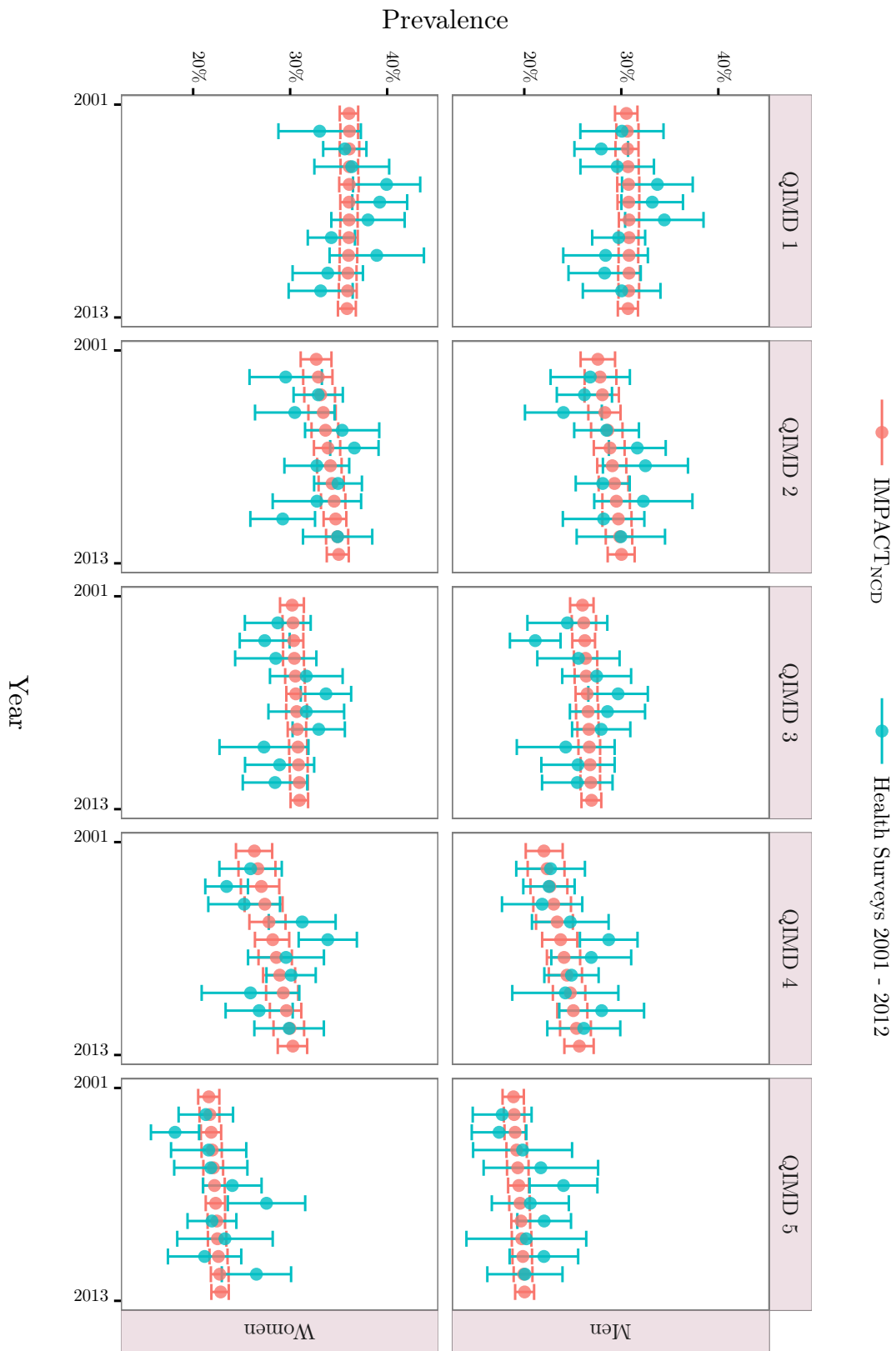
Figure B.26: Prevalence of consumption of five or more fruit and vegetable portions per day for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
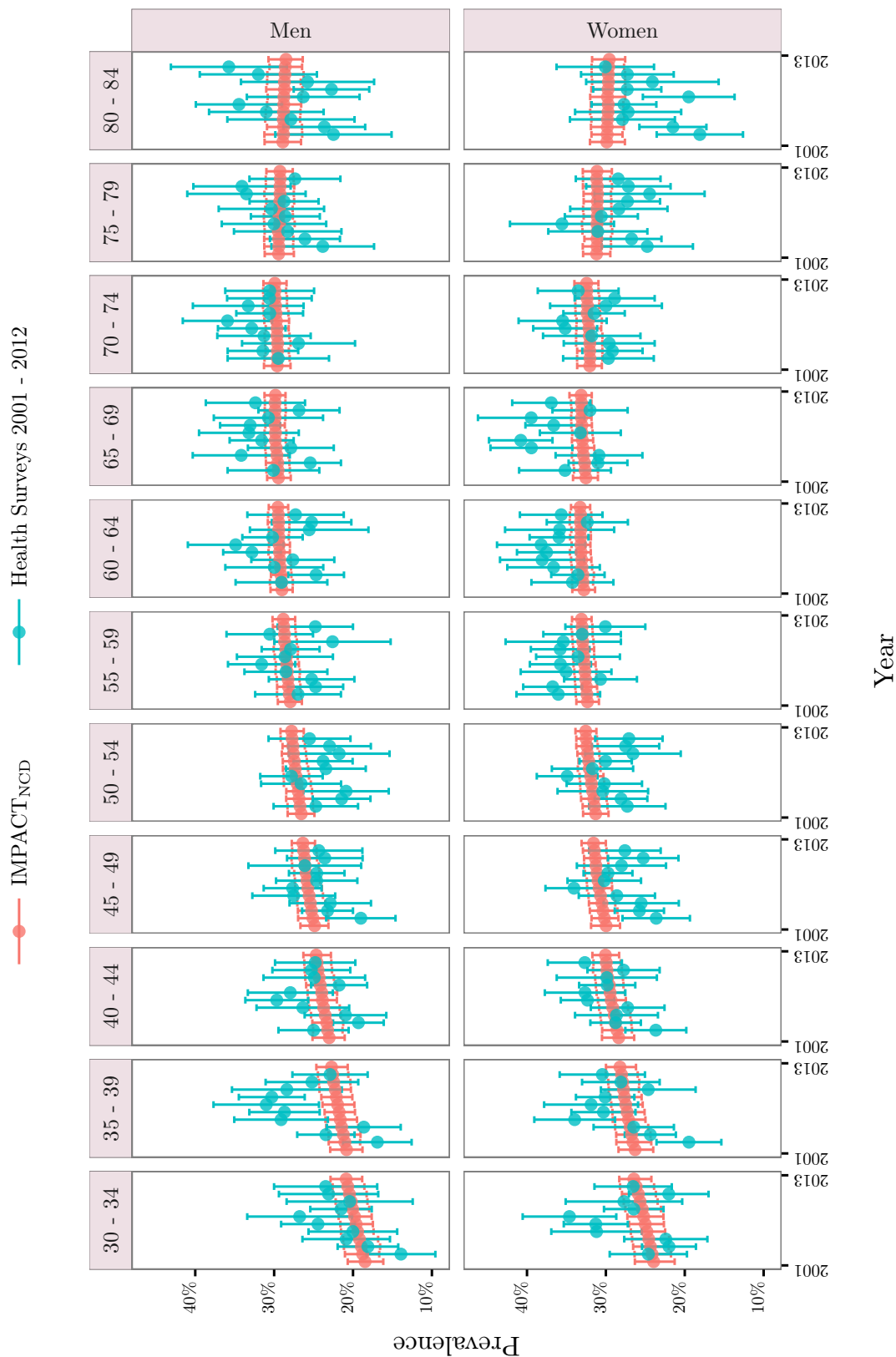
Figure B.27: Prevalence of consumption of five or more fruit and vegetable portions per day for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.
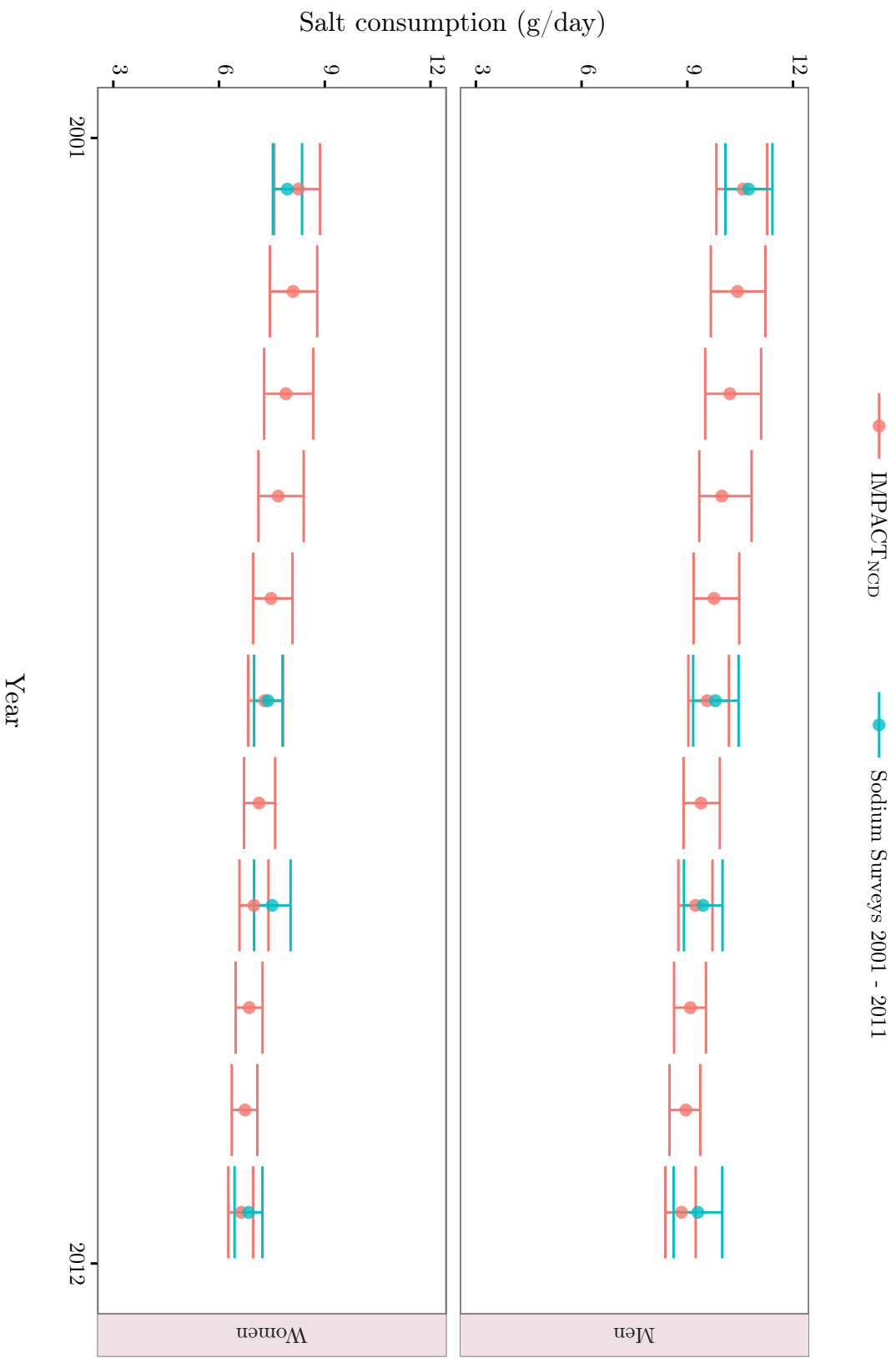
Figure B.28: Mean salt consumption for ages 19 – 64 between years 2001 and 2011. Observed in the population through surveys using 24 h urine collections[68, 240–242] versus IMPACT$_{NCD}$ synthetic population estimates. Error bars depict 95 % confidence intervals of the mean.
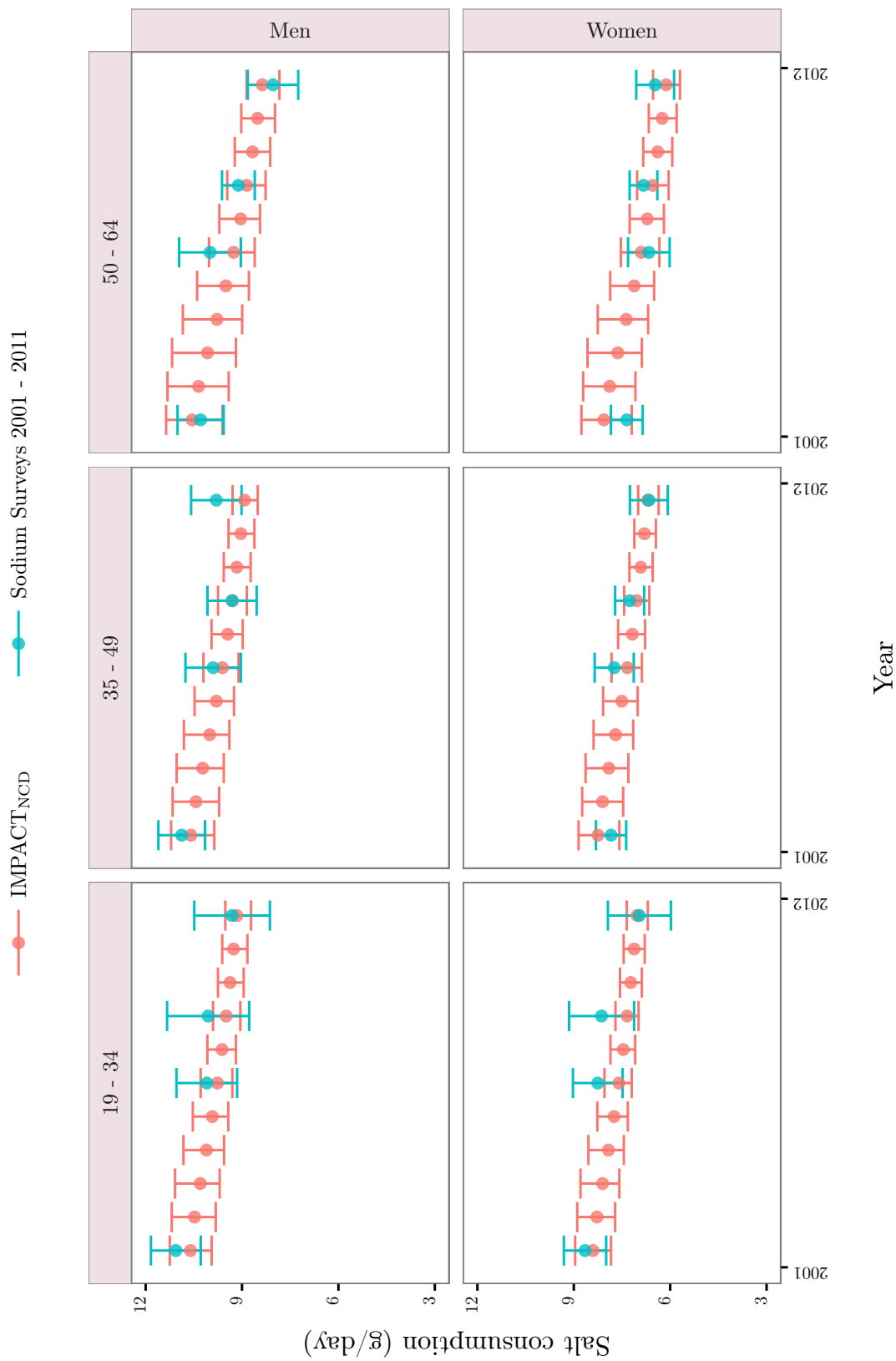
Figure B.29: Mean salt consumption between years 2001 and 2011, by age group. Observed in the population through surveys using 24 h urine collections[68, 240–242] versus IMPACT$_{NCD}$ synthetic population estimates. Error bars depict 95 % confidence intervals of the mean.

Figure B.30: Prevalence of five or more active days per week for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.31: Prevalence of five or more active days per week for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
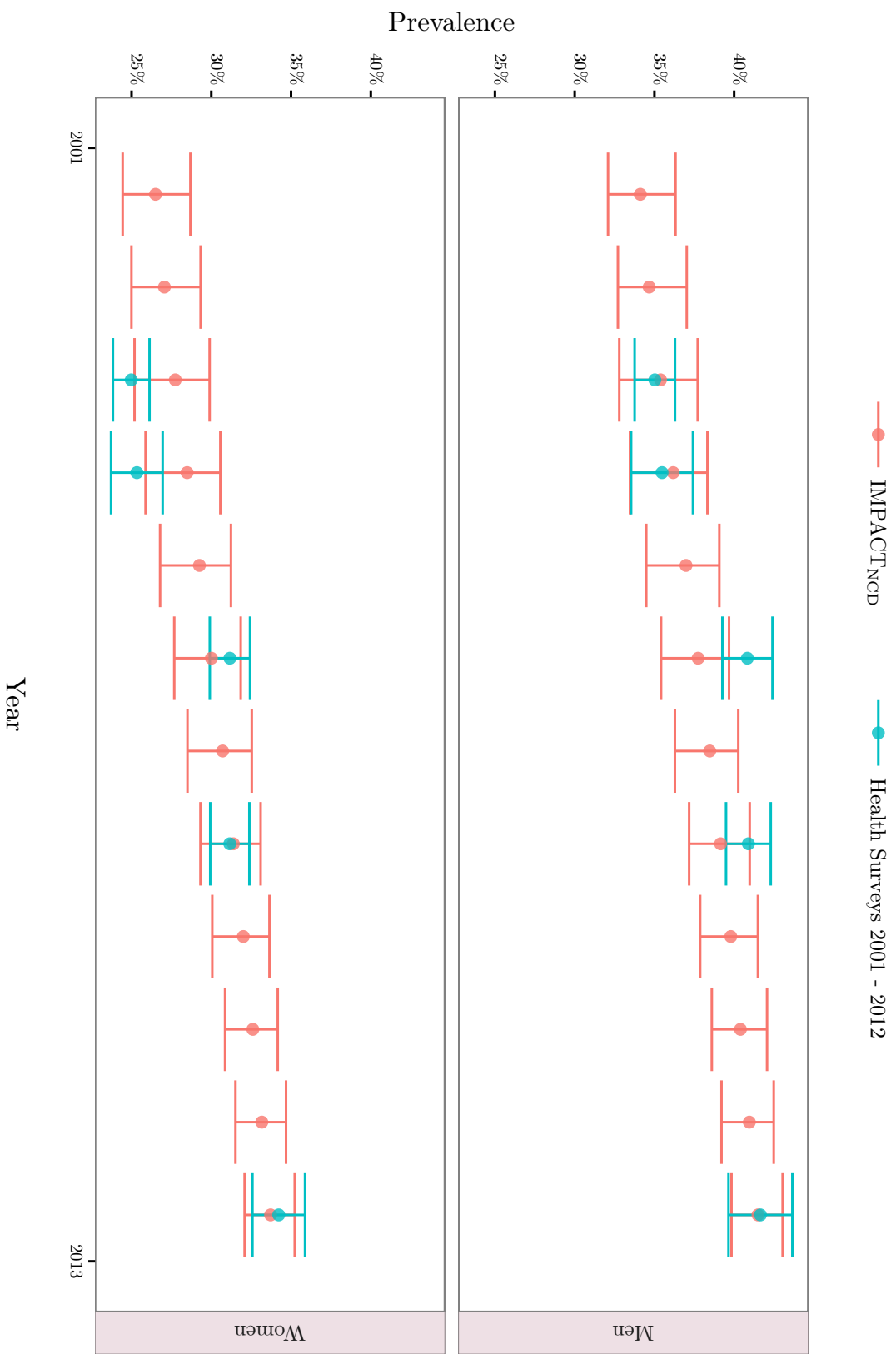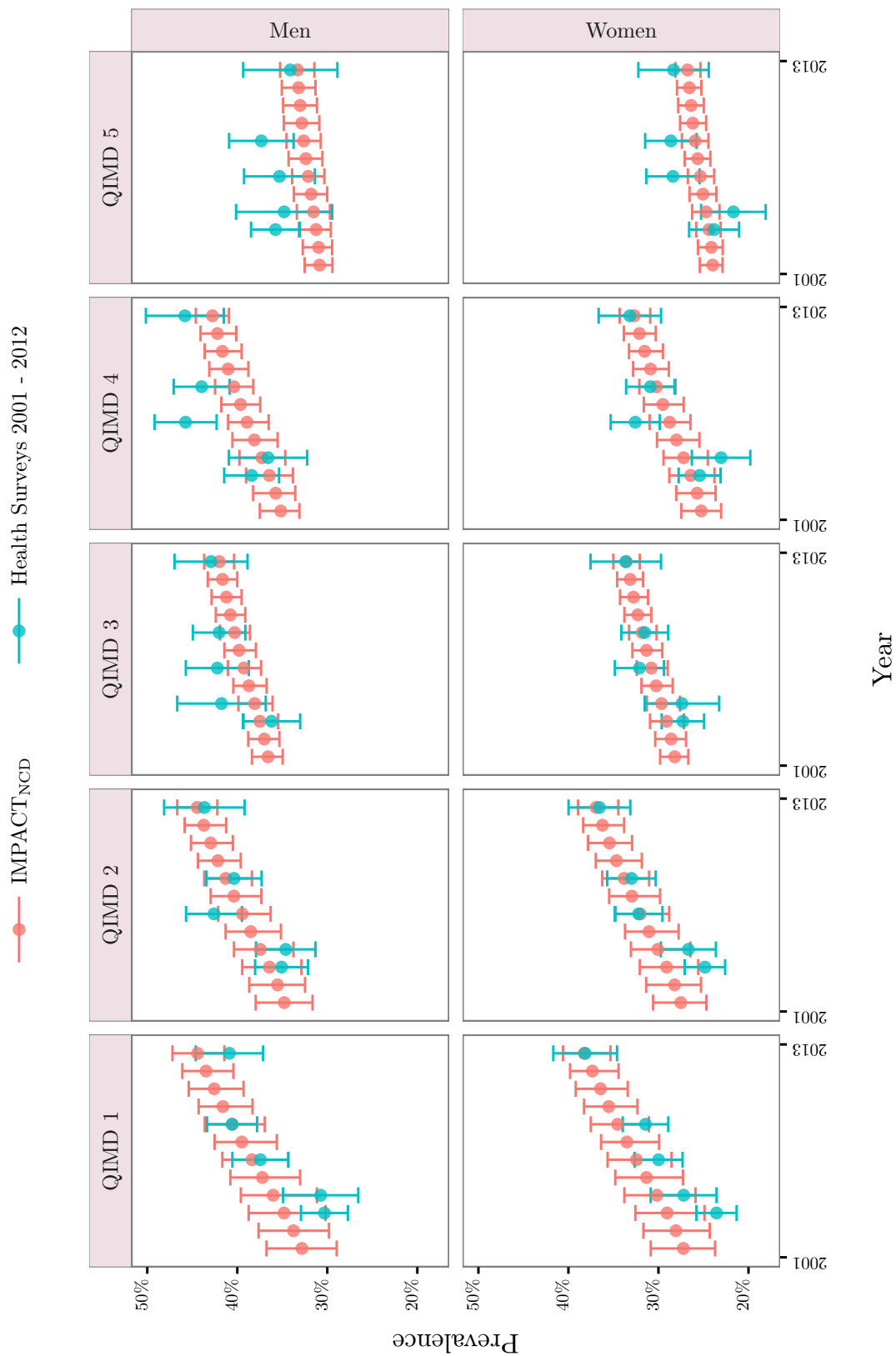
Figure B.32: Prevalence of five or more active days per week for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.33: Mean body mass index for ages 30 − 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.34: Mean body mass index for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
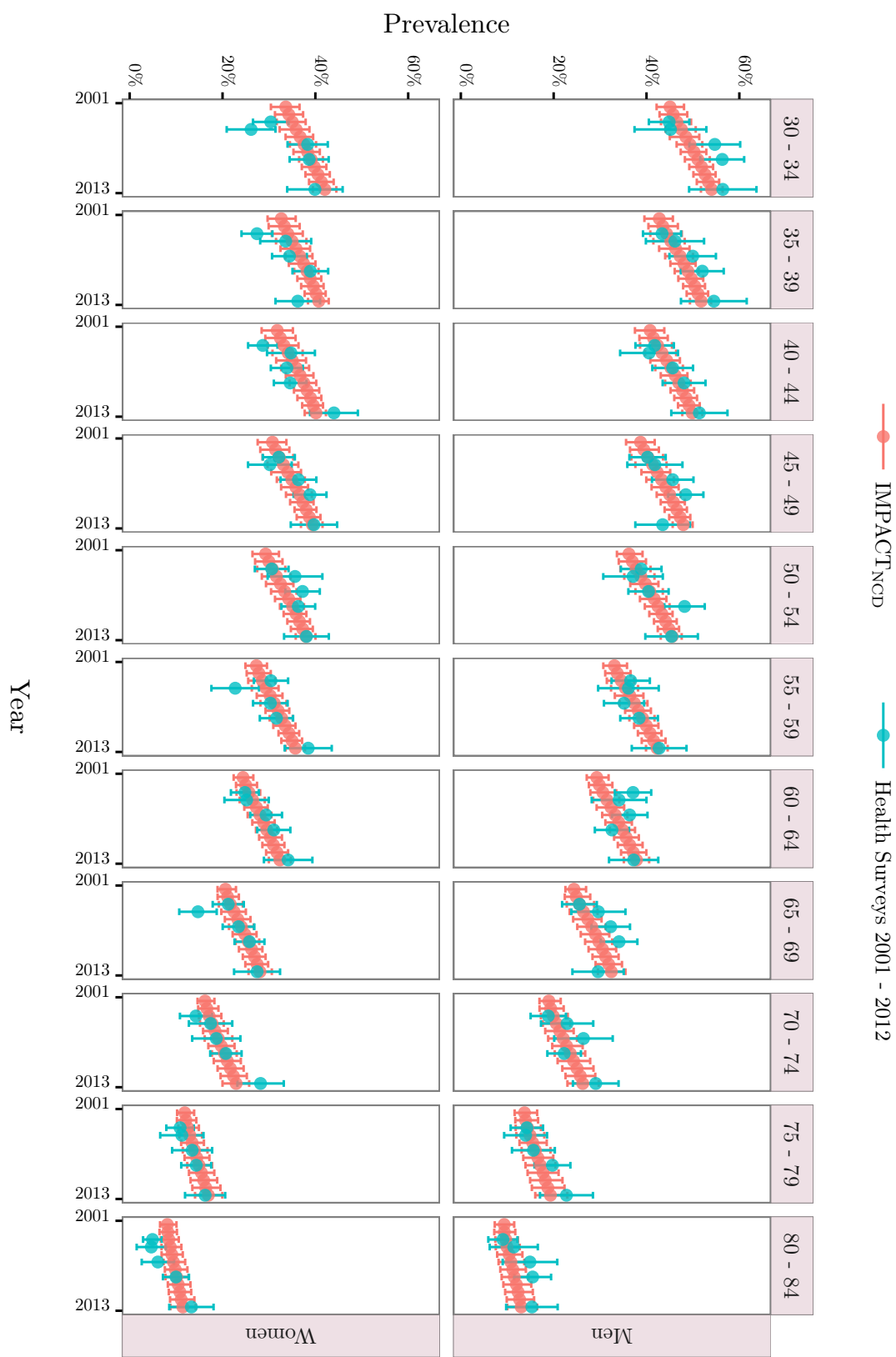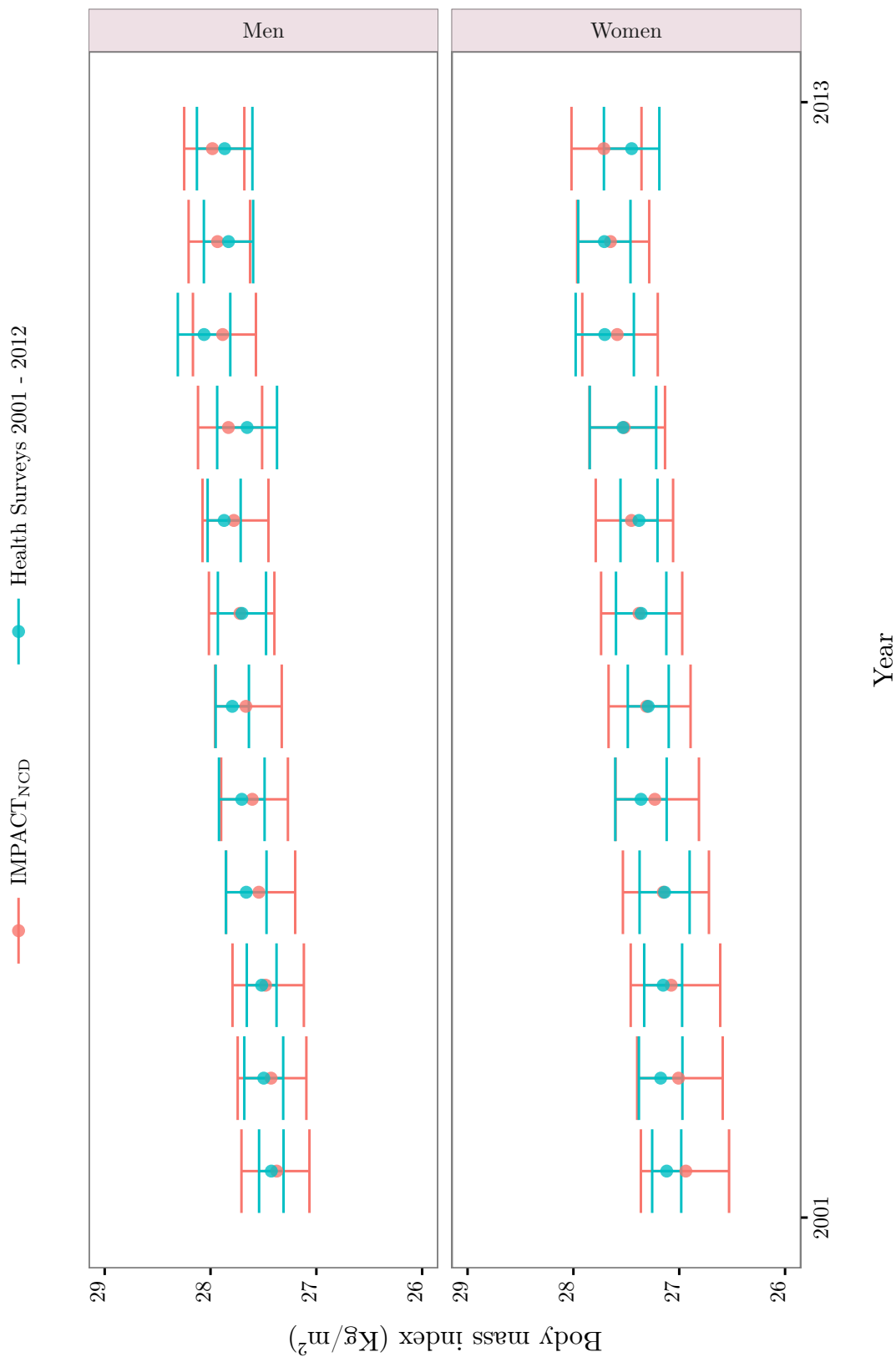
Figure B.35: Mean body mass index for ages 30 − 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.36: Diabetes mellitus prevalence for ages $30 - 84$ between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95% confidence intervals of the mean.
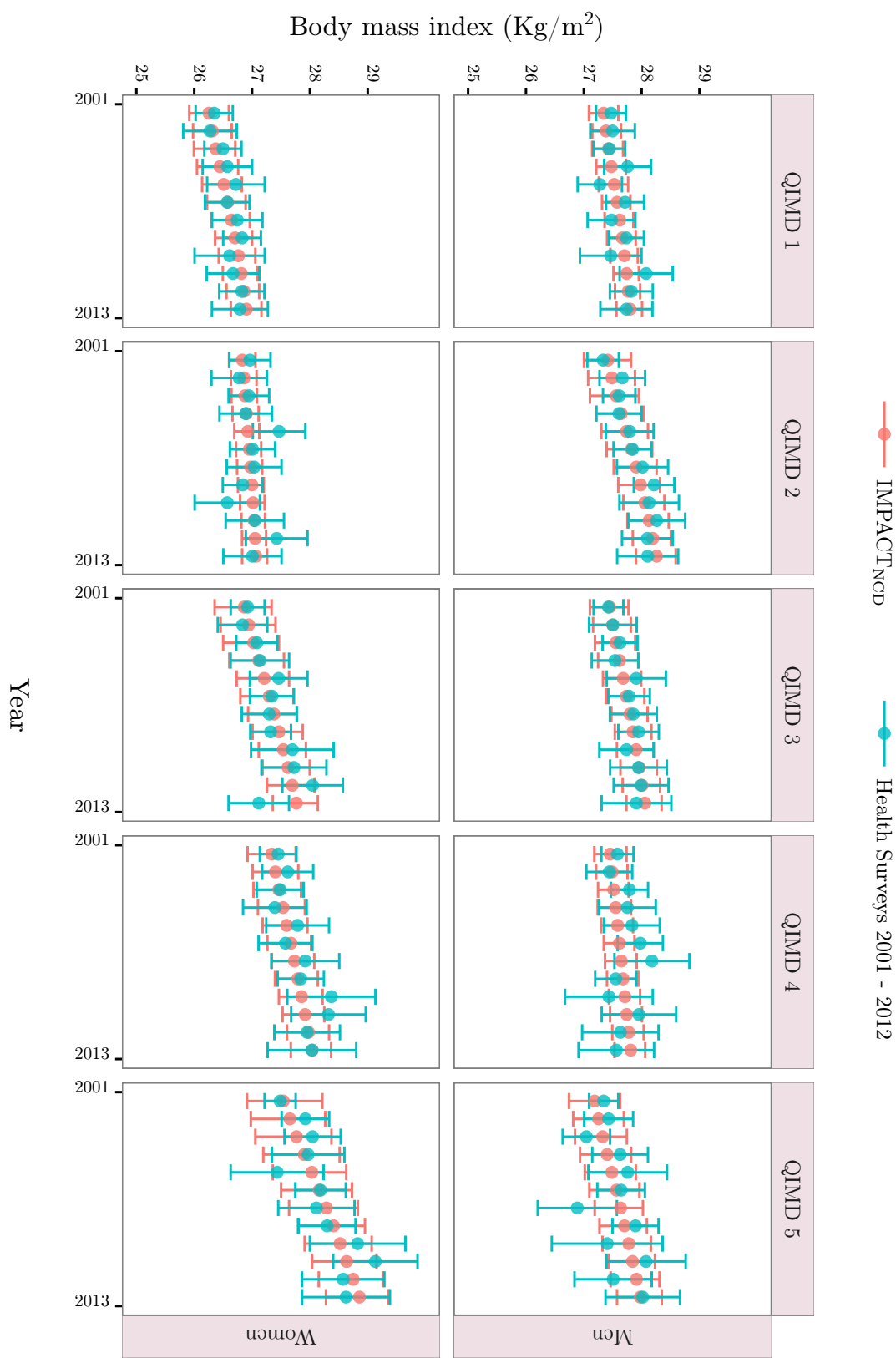
Figure B.37: Diabetes mellitus prevalence for ages 30 − 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
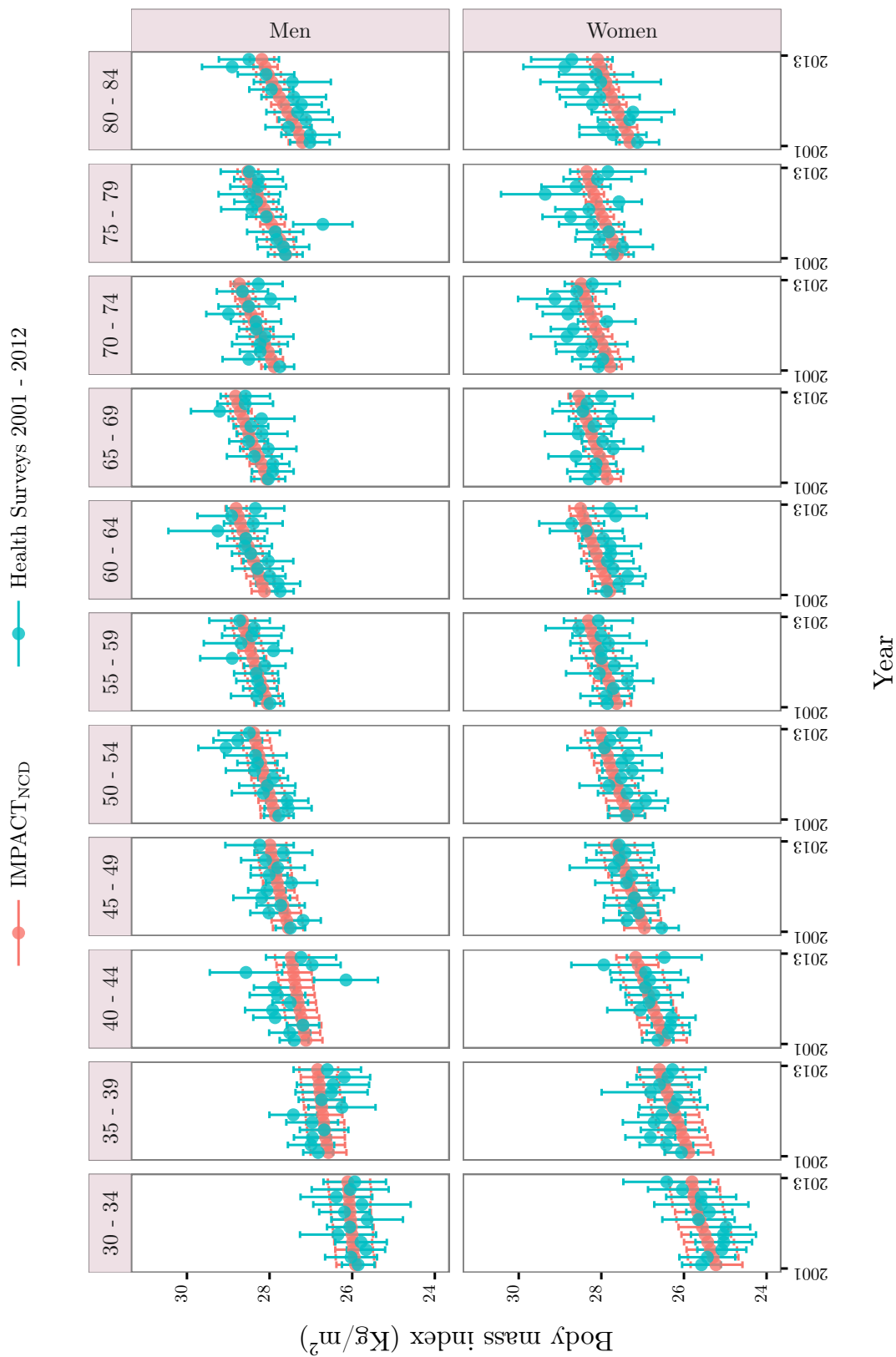
Figure B.38: Diabetes mellitus prevalence for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{\text{NCD}}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.

Figure B.39: Mean total cholesterol for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95 % confidence intervals of the mean.
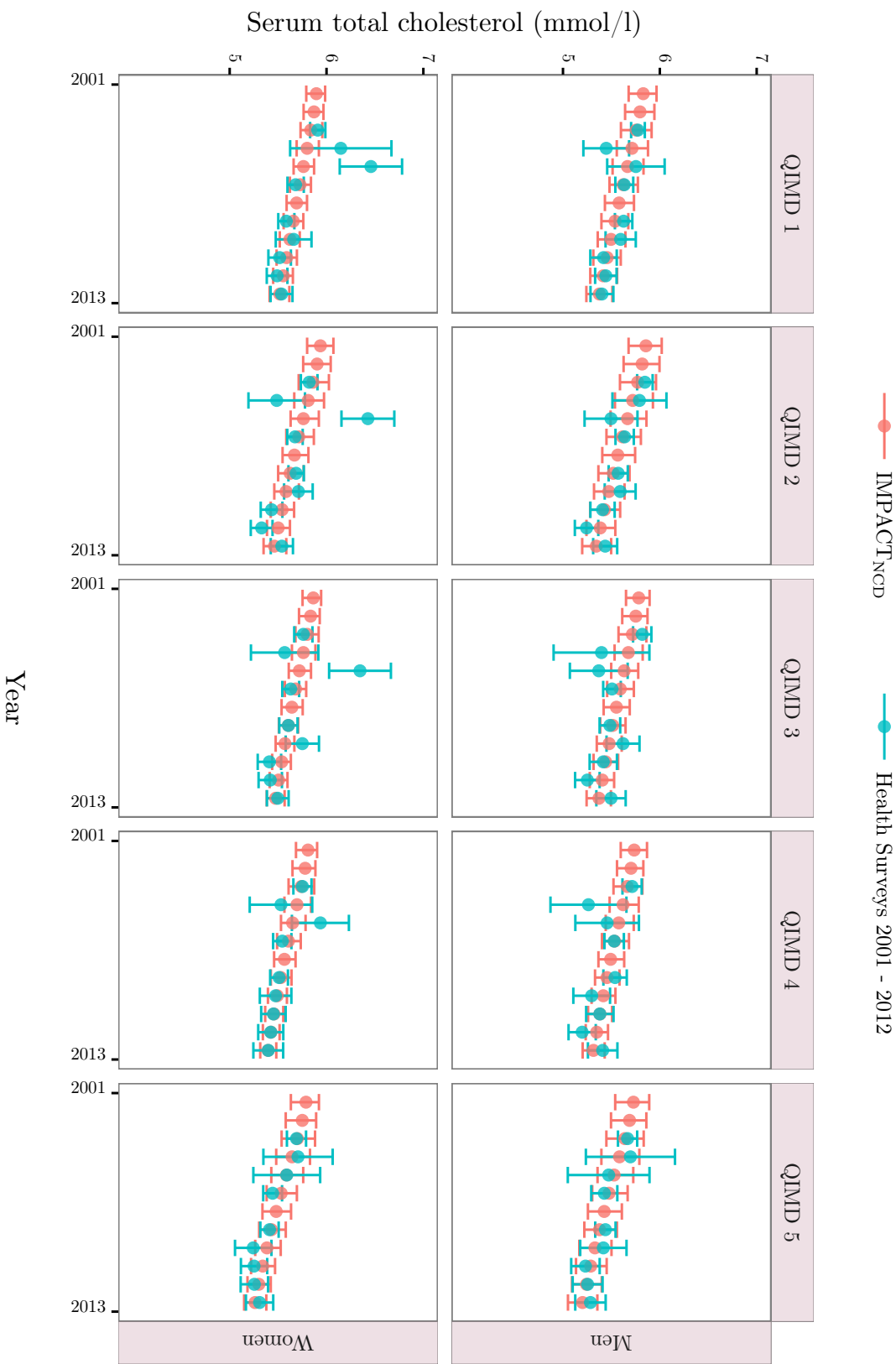
Figure B.40: Mean total cholesterol for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
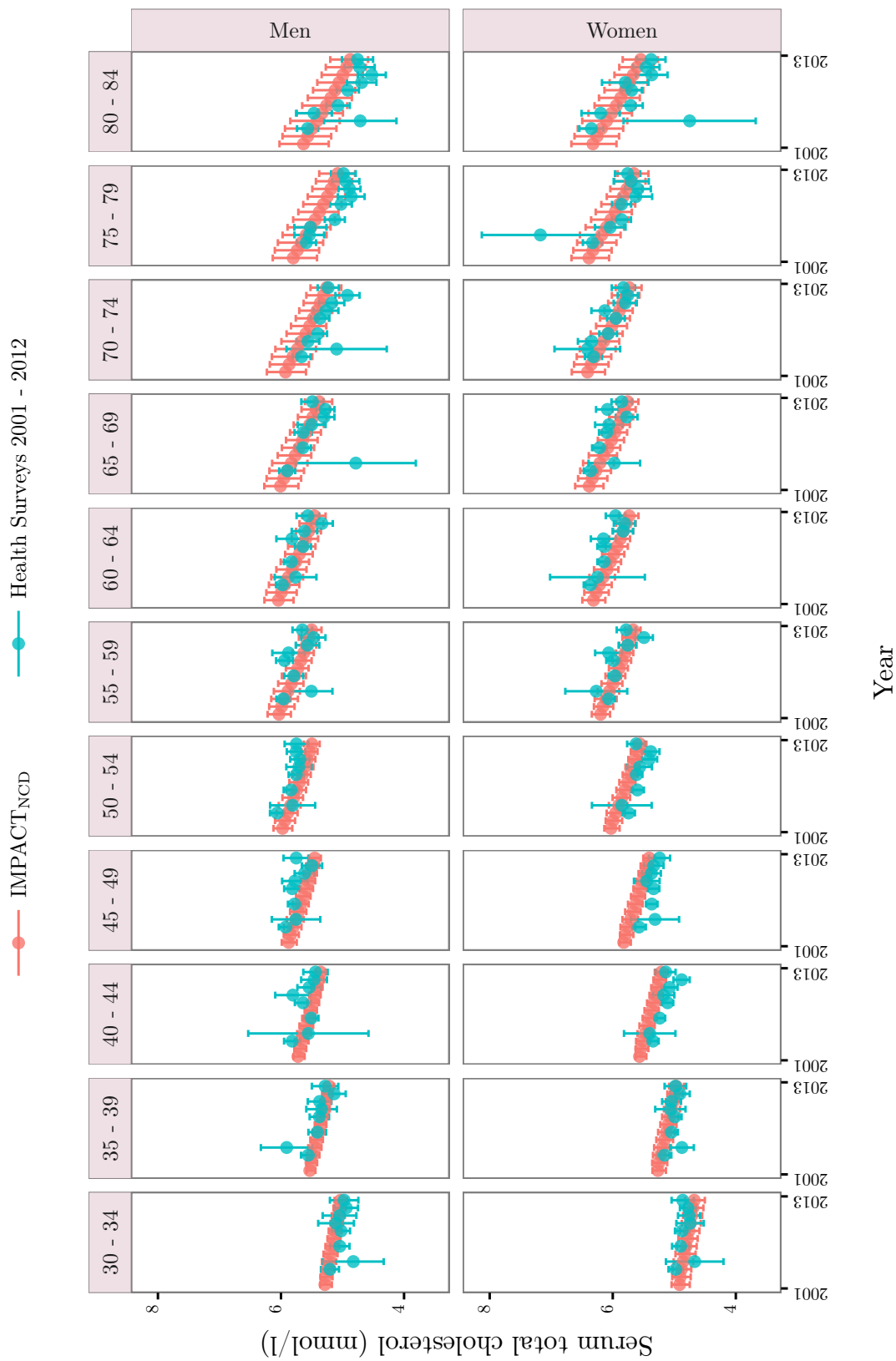
Figure B.41: Mean total cholesterol for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.
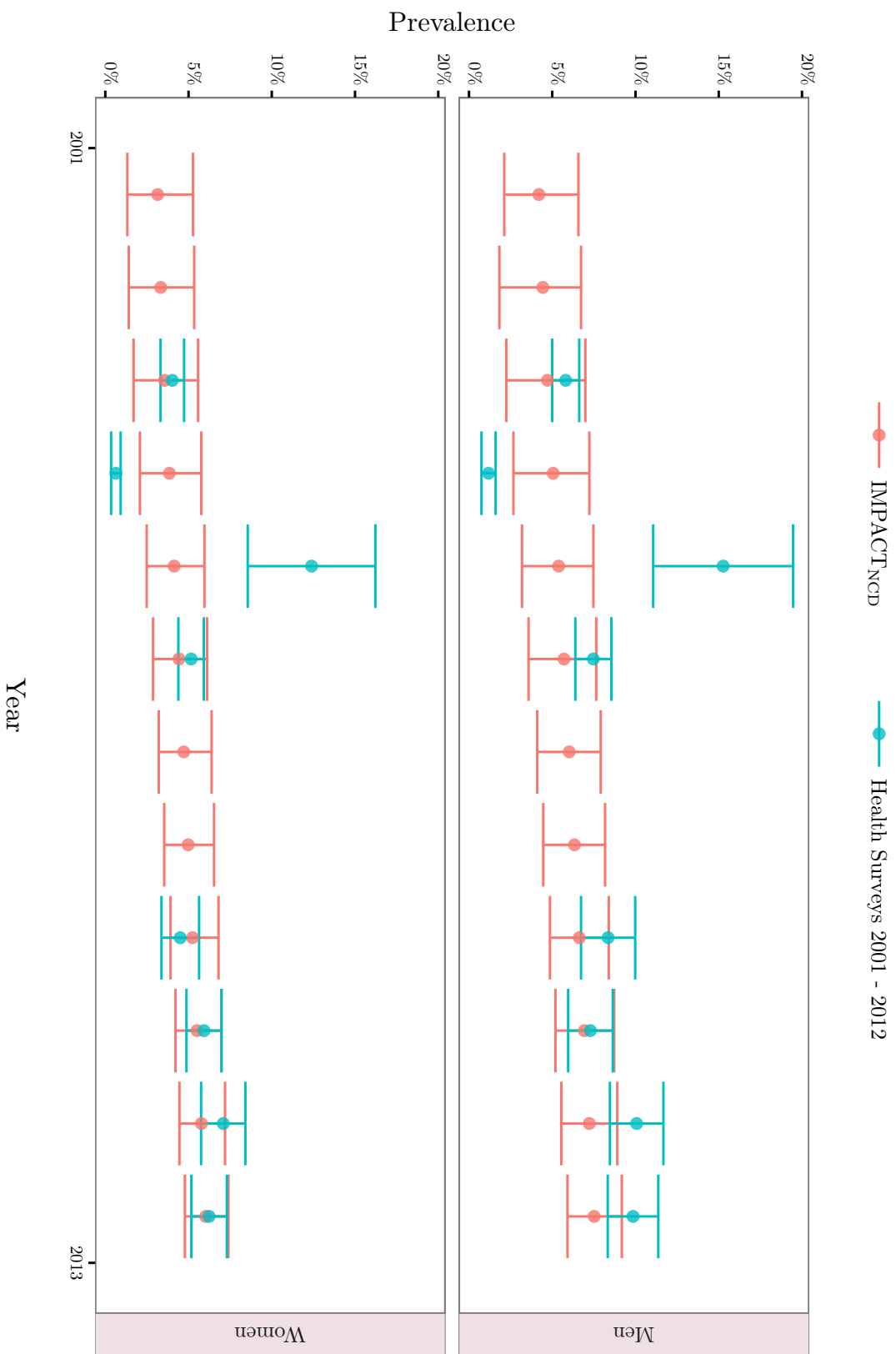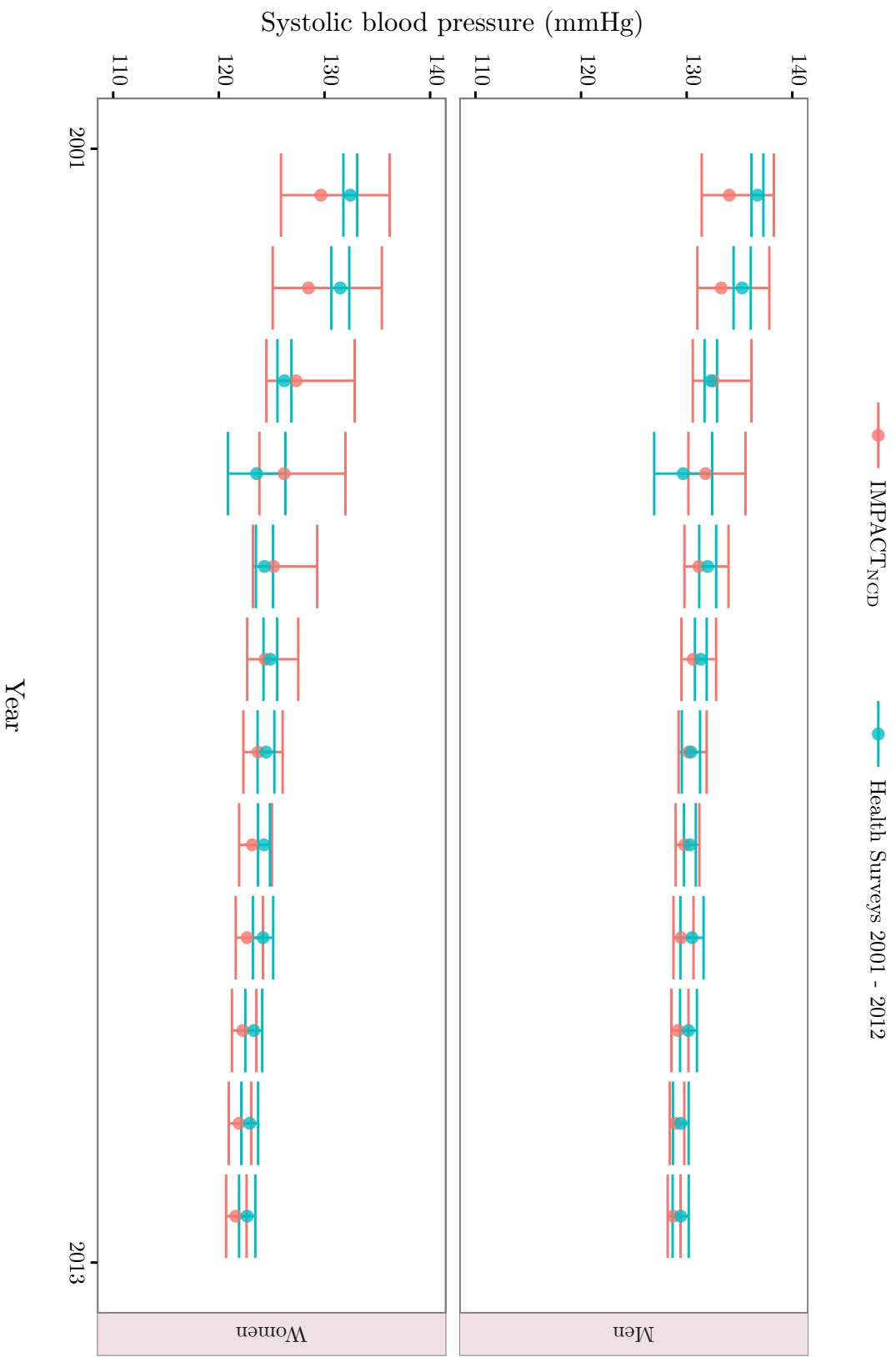
Figure B.42: Mean systolic blood pressure for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT_NCD synthetic population estimates, stratified by sex. Error bars depict 95 % confidence intervals of the mean.
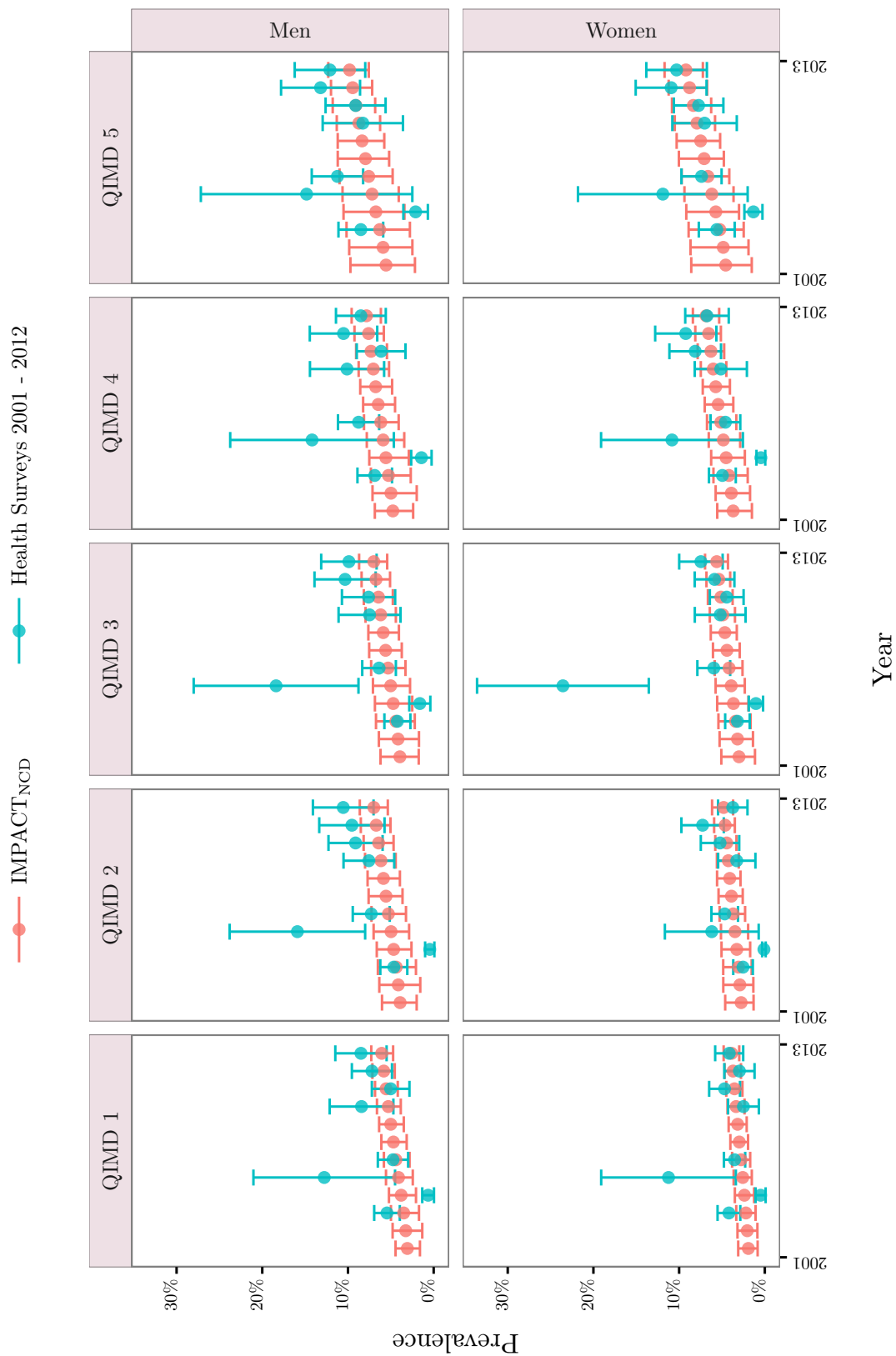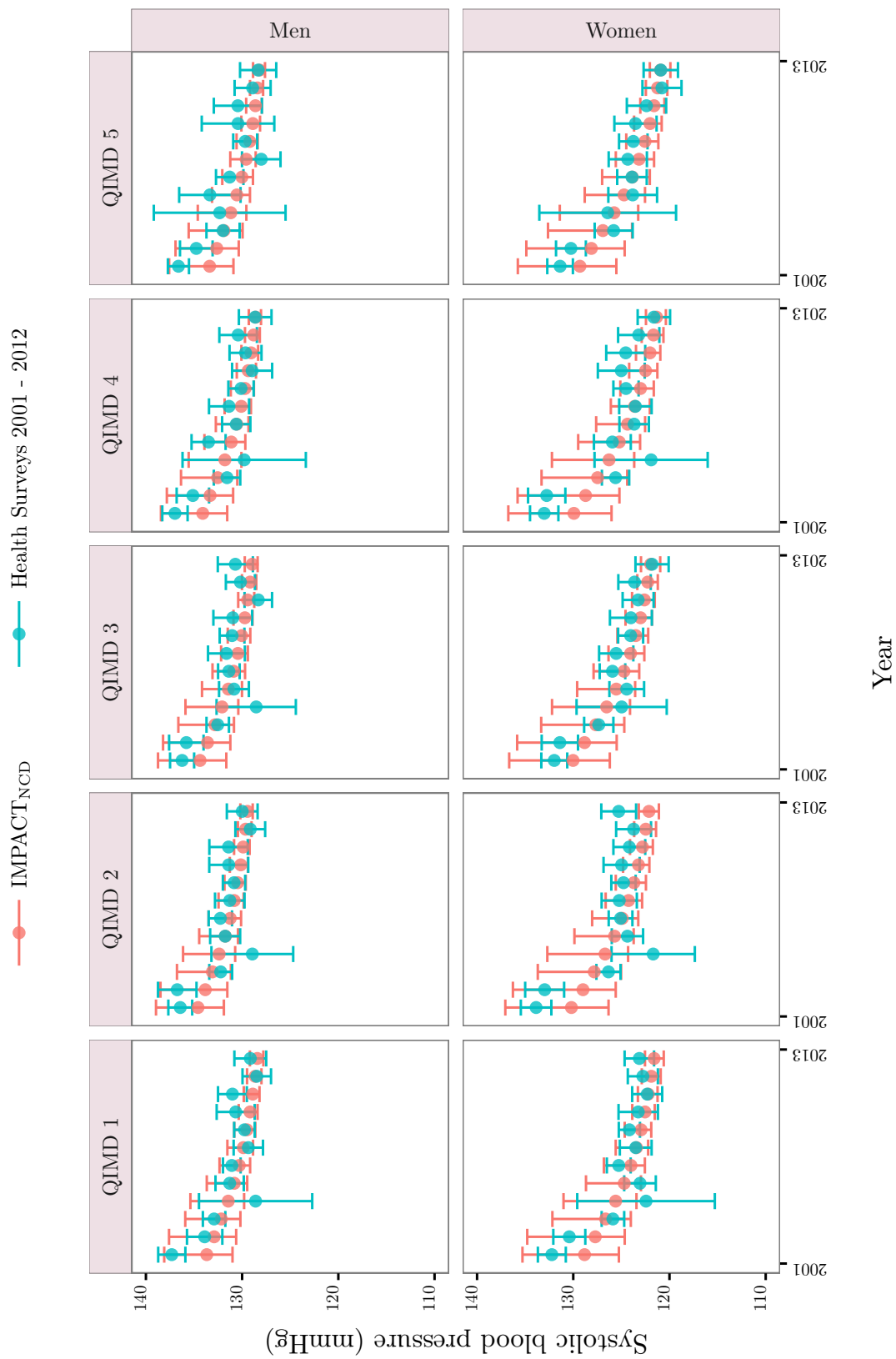
Figure B.43: Mean systolic blood pressure for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by quintile groups of Index of Multiple Deprivation (QIMD, 1 = least deprived, 5 = most deprived) and sex. Error bars depict 95 % confidence intervals of the mean.
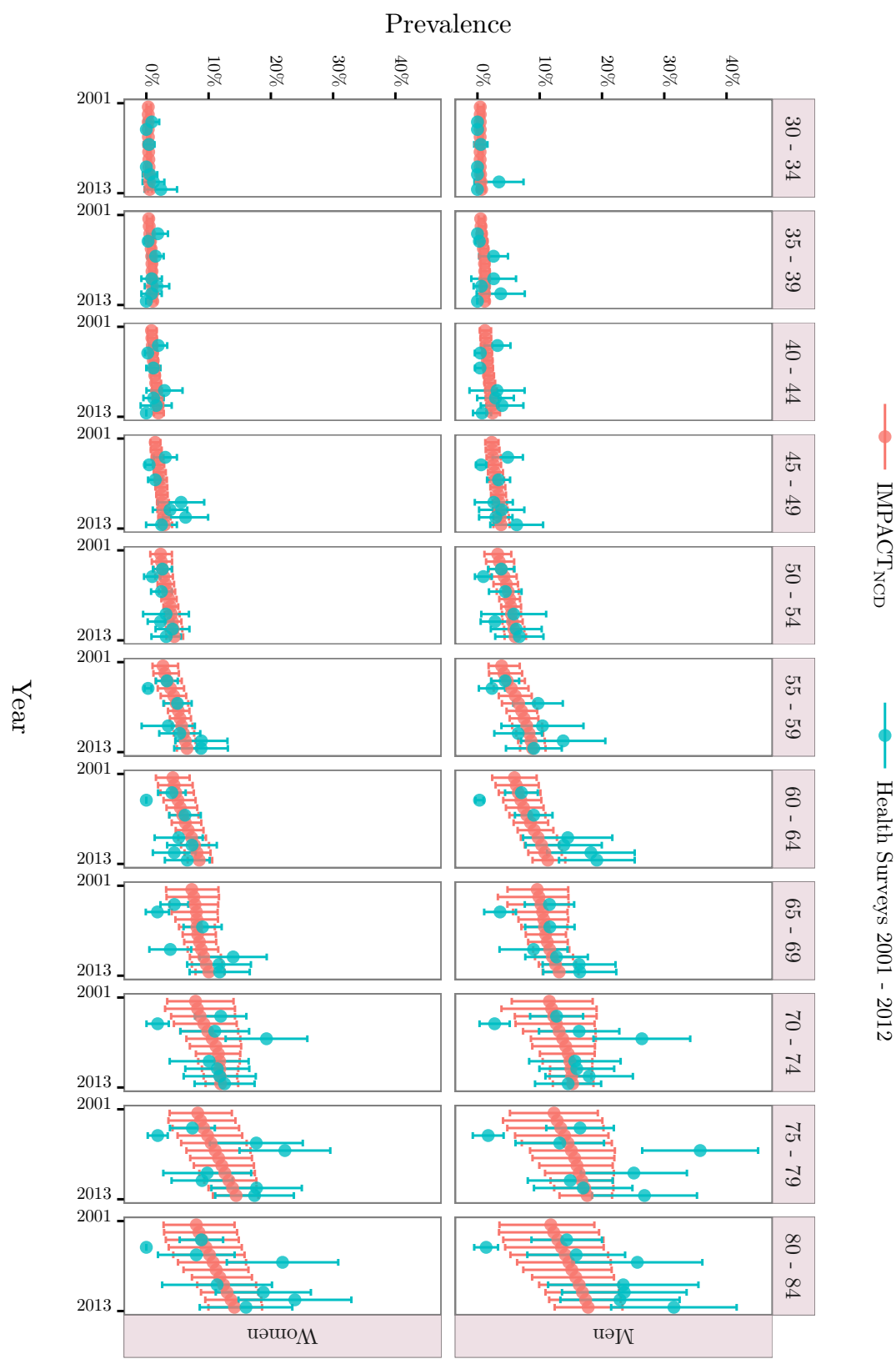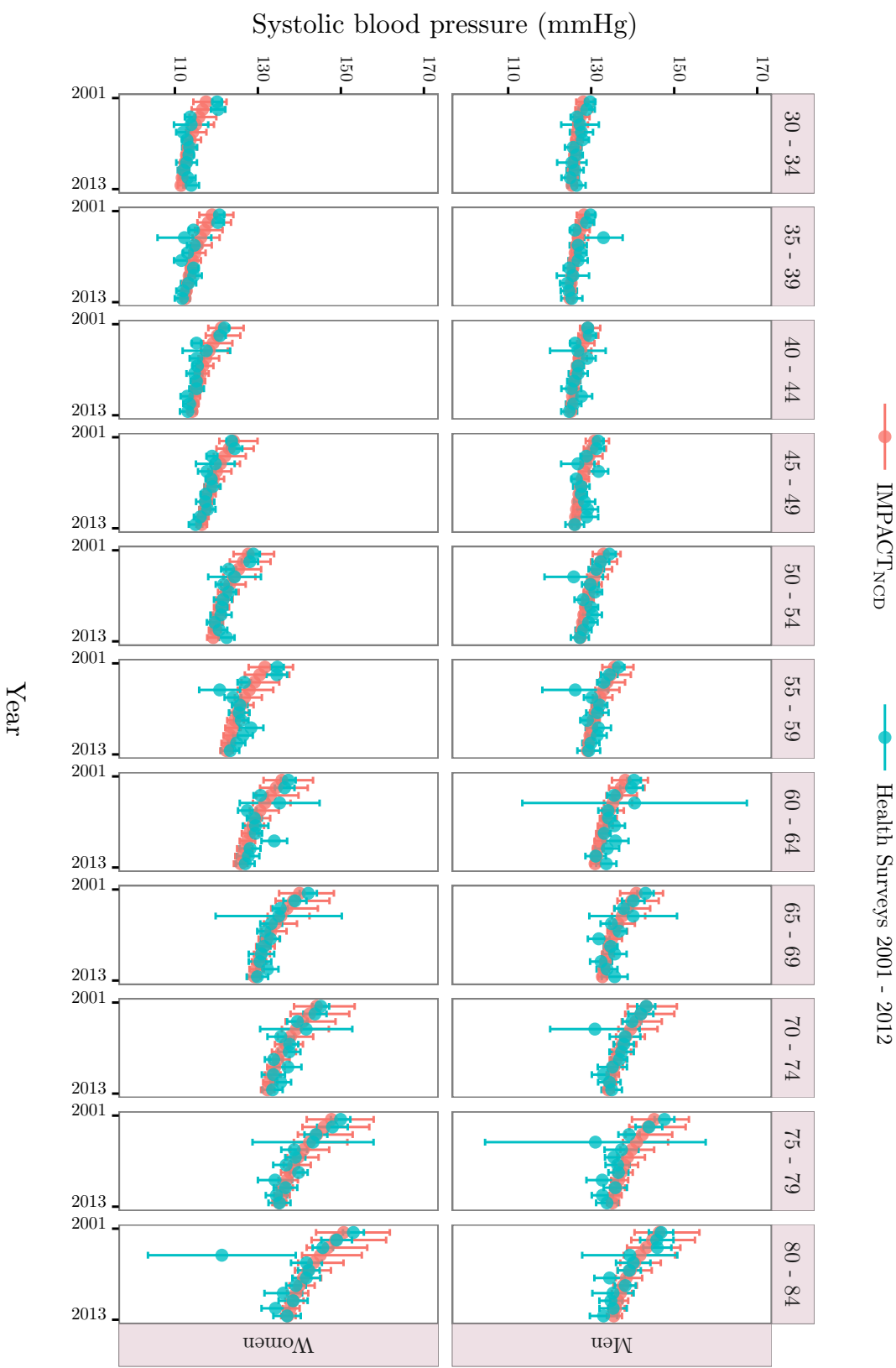
Figure B.44: Mean systolic blood pressure for ages 30 – 84 between years 2001 and 2012. Observed in the population through Health Survey for England versus IMPACT_NCD synthetic population estimates, stratified by age group and sex. Error bars depict 95 % confidence intervals of the mean.
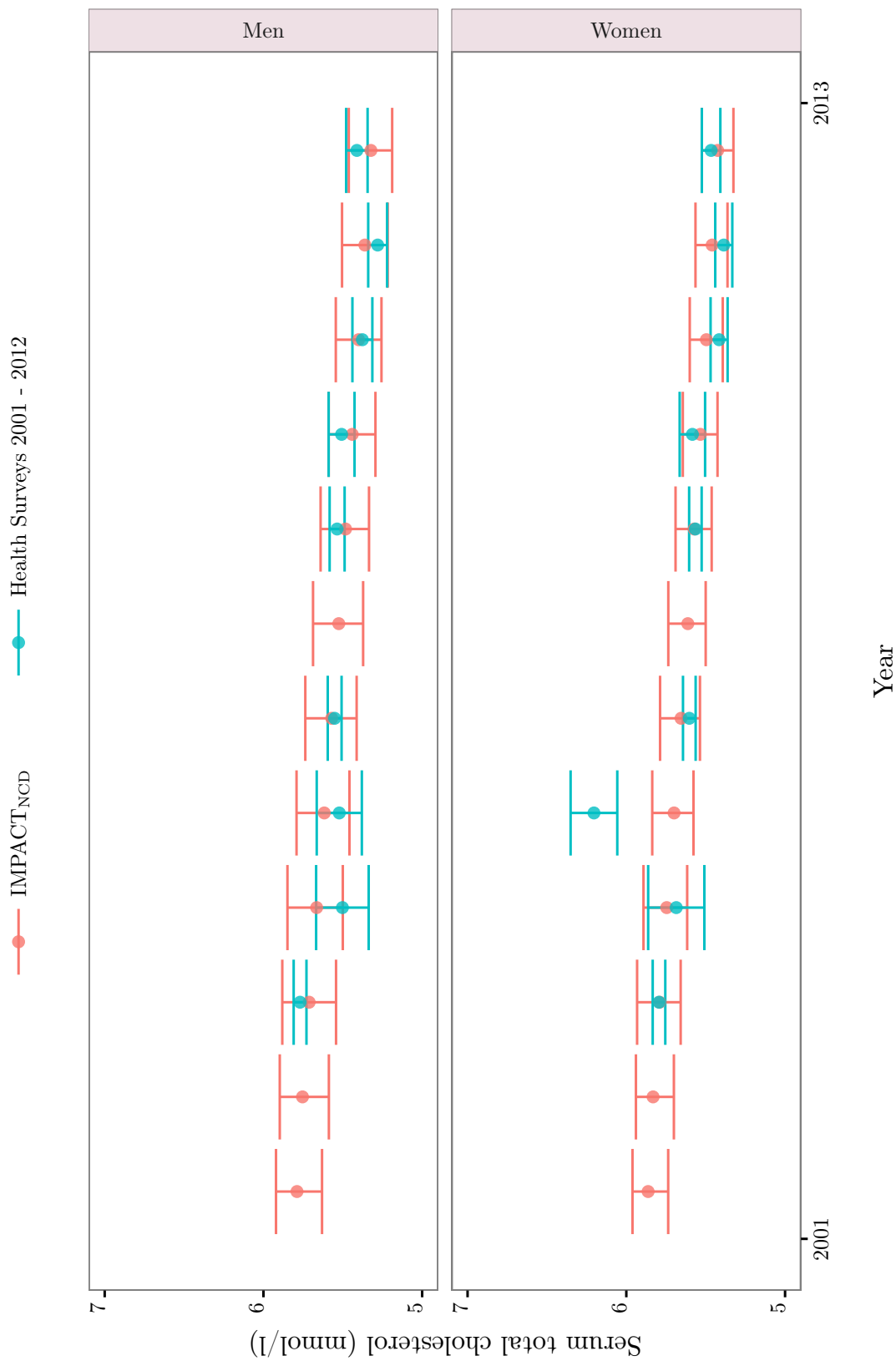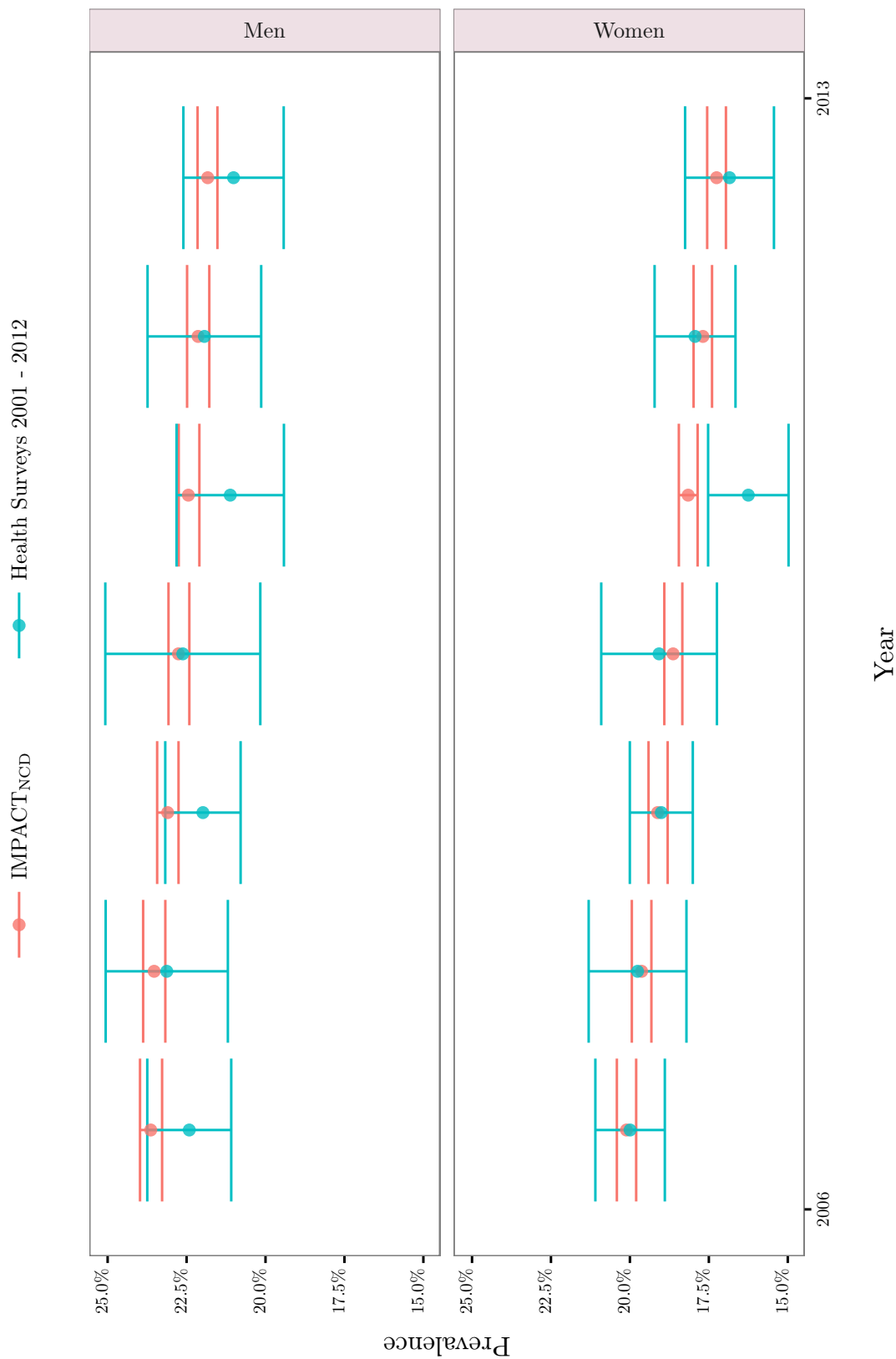
Figure B.45: Smoking prevalence for ages 30 – 84 between years 2006 and 2012. Observed in the population through Health Survey for England versus IMPACT$_{NCD}$ synthetic population estimates, stratified by sex. Error bars depict 95% confidence intervals of the mean. In comparison to figure B.16 on page 214 this graph was plotted using smoking prevalence with no lag time.

RESULTS APPENDIX

C.1 SUPPORTING THE ASSUMPTION OF NO STATIN EFFECT IN 1991-92, PERTAINING CHAPTER 4

In 1991-92 statins were not widely used, and only reserved for individuals with very high cholesterol. This suggested from the HSE1991-92 data; utilisation of any lipid lowering medications including statins was 0.5% (95% CI: 0.3% to 1.0%) and the mean total cholesterol of participants on this medication was 6.46 mmol/l (95% CI: 6.04 mmol/l to 6.87 mmol/l). Moreover, statins available at that time were less effective and generally were prescribed in smaller strengths than today. Therefore, their estimated effectiveness (E_w) for 1991-92 would be much lower than the E_w for 2011-12. Even if all the lipid lowering medication users in 1991-92 were on statins and statins effectiveness (E_w) was as high as I estimated for 2011-12, the mean total cholesterol of the population for 1991-92 with the effect of statin removed, would be 5.87 mmol/l (95% CI: 5.83 mmol/l to 5.91 mmol/l). Not much different from the observed one of 5.86 mmol/l (95% CI: 5.82 mmol/l to 5.90 mmol/l). Therefore, I consider the bias from my decision to ignore any possible statin effect in 1991-92 negligible.

C.2 EFFECT OF STATINS ON REDUCTION OF TOTAL CHOLESTEROL, PERTAINING CHAPTER 4

Table C.1: Final percentage reductions of each specific statin and strength that were used for the estimation of the weighted average E_w.

Chemical name	Strength in mg	Total cholesterol reduction (95% confidence intervals)	Weights (for the weighted mean E_w)
Atorvastatin	10	27.3% (24.9% - 30.2%)	0.1046
Atorvastatin	20	32.7% (30.1% - 35.7%)	0.0361
Atorvastatin	30	35.8% (34.8% - 36.7%)*	0.0005
Atorvastatin	40	38.4% (34.6% - 42.3%)	0.0350
Atorvastatin	60	41.0% (39.7% - 42.3%)*	0.0005
Atorvastatin	80	42.8% (37.4% - 48.0%)	0.0118
Fluvastatin Sodium	20	16.4% (14.6% - 18.4%)	0.0006
Fluvastatin Sodium	40	20.7% (19.0% - 22.5%)	0.0173

...continued

Chemical name	Strength in mg	Total cholesterol reduction (95% confidence intervals)	Weights (for the weighted mean E_w)
Fluvastatin Sodium	80	23.3% (20.6% - 25.9%)	0.0163
Pravastatin Sodium	5	10.4% (0.7% - 20.2%)	0.0001
Pravastatin Sodium	10	14.5% (12.5% - 16.2%)	0.0038
Pravastatin Sodium	20	17.7% (16.9% - 18.9%)	0.0111
Pravastatin Sodium	40	22.0% (20.7% - 23.0%)	0.0106
Rosuvastatin Calcium	5	25.9% (24.7% - 27.6%)	0.0114
Rosuvastatin Calcium	10	29.0% (27.8% - 30.6%)	0.0214
Rosuvastatin Calcium	20	32.1% (30.9% - 33.6%)	0.0042
Rosuvastatin Calcium	40	35.2% (34.0% - 36.6%)	0.0012
Simvastatin	10	20.1% (18.9% - 21.7%)	0.0477
Simvastatin	20	23.5% (22.4% - 25.0%)	0.4261
Simvastatin	25	24.6% (24.3% - 25.0%)*	0.0001
Simvastatin	40	27.0% (25.2% - 28.9%)	0.2339
Simvastatin	80	30.4% (29.6% - 31.3%)	0.0045
Simvastatin & Ezetimibe	20	23.5% (22.4% - 25.0%)	0.0003
Simvastatin & Ezetimibe	40	27.0% (25.2% - 28.9%)	0.0008
Simvastatin & Ezetimibe	80	30.4% (29.6% - 31.3%)	0.0001

* Values derived from log-linear regression with total cholesterol reduction as the dependent variable and the natural logarithm of strength as the independent one. The model was weighted against the inverse variance of the cholesterol reduction (not presented in this table).

C.3   EXTRA SCENARIO SPECIFICATIONS, PERTAINING CHAPTER 5

In the following paragraphs, I highlight some details of the scenarios that I used in the main paper. They are meant to be read in conjunction with the scenario description in the main text (section 5.2.1 on page 122), and the methods section (section 2.5 on page 61).

UNIVERSAL SCREENING: This was a typical targeted intervention, so this scenario was built with the second approach in section 2.5 on page 61. The high-risk synthetic individuals eligible for treatment were selected based on the QRISK2 score.[336] The score requires extra information about the synthetic individual that was not originally modelled and at the current stage is used exclusively for the calculation of the QRISK2 score. This includes information about ethnicity, specific type of diabetes mellitus (I or II), family history of CVD, chronic kidney disease (stage 4 or 5), atrial fibrillation, rheumatoid arthritis,

and the total cholesterol to high density lipoprotein ratio. To model these extra attributes for the synthetic individuals I fitted appropriate multinomial, logistic, or generalised linear regression models to HSE data, and then I used the models to predict synthetic individuals' status. Exceptions, to this approach were type I diabetes mellitus and rheumatoid arthritis prevalence. I assumed a prevalence of 0.5 % for diabetes mellitus type I and I extracted age and sex specific rheumatoid arthritis prevalence from published data.[471]

To simulate ethnicity of synthetic individuals, a multinomial model was fitted to HSE data with 5-year age group, sex and QIMD as the independent variables. To simulate family history of CVD, a logistic regression model was fitted in HSE2006 data that contained this information, with age and QIMD as the independent variables. For the prevalence of atrial fibrillation, a logistic regression model was fitted in HSE2011 data that contained this information, with age, QIMD, and smoking status as the independent variables. To model the prevalence of chronic kidney disease a logistic regression model with age, sex, and QIMD as independent variables was fitted to HSE2010 data. Finally, for the total cholesterol to high density lipoprotein ratio a regression was fitted to HSE data, with total cholesterol, age, sex, QIMD, BMI, physical activity, and smoking status as the independent variables.[55]

To estimate the individualised effectiveness of Atorvastatin, I used the formula:

$$Individualised\ effectiveness =$$
$$= Effectiveness * Prescription * Persistence * Adherence \quad \text{(C.1)}$$

Where *Effectiveness* for Atorvastatin 20 mg was extracted from table C.1 on page 245, as I estimated in section 4.2.5.1 on page 106, *Prescription* is a binary variable (whether Atorvastatin was prescribed (1), or not (0)), *Persistence* is a binary variable (whether the synthetic individual continue with the medication (1), or not (0)), and *Adherence* is a value between 0 and 1, modelling the proportion of daily dose taken. For these variables, values where drawn from distributions (table A.1 on page 187).

I followed a similar approach for antihypertensive medication. Given the numerous antihypertensive treatment combinations, I assumed that medication could fully control hypertension for all synthetic individuals down to a target of 115 mmHg of SBP if prescription, persistence, and adherence were optimal. I applied the same approach as above to adjust treatment effectiveness to prescription, persistence, and adherence.

Information regarding medication prescription after a Health Check was extracted from Forster *et al.*[376] This study was conducted while the recommendation for primary prevention statin prescription was based on 20 % risk for a CVD event in 10 years. Yet, statin prescription was low in this group and statins were prescribed to individuals with lower than 20 % risk. I chose to inflate the reported from Forster *et al.* prescription rate for participants with a risk between 10 % and 20 %. This was made to reflect the recent change in recommendation about statin prescription for primary prevention, which lower the

---

55 The R objects for the models are available at https://github.com/ChristK/IMPACTncd/blob/CVD-policy-opt ions/Lagtimes/origin.multinom.rda, https://github.com/ChristK/IMPACTncd/blob/CVD-policy-options/Lagti mes/famcvd.svylr.rda, https://github.com/ChristK/IMPACTncd/blob/CVD-policy-options/Lagtimes/af.svylr .rda, https://github.com/ChristK/IMPACTncd/blob/CVD-policy-options/Lagtimes/kiddiag.svylr.rda, https:// github.com/ChristK/IMPACTncd/blob/CVD-policy-options/Lagtimes/tctohdl.svylm.rda.

threshold from 20 % to 10 % risk for a CVD event in 10 years. I avoided making it equal to the prescription rate of those with a risk higher than 20 %, based on finding from Usher-Smith *et al.* that reported reluctant statin uptake to the newly eligible population.[472]

In one of the simulated scenarios for sensitivity analysis, I assumed a treatment threshold of 20 % risk for a CVD event in 10 years. For this scenario, I used prescription rates as reported from Forster *et al.* for the participants with a risk higher than 20 %.[376] Yet, I also allowed synthetic participants to be prescribed medication with a risk higher than 10 % as was reported in the study. The justification was that despite the recommended 20 % threshold to offer treatment to high-risk individuals, when the study from Forster *et al.* was contacted, participants with lower risk were still prescribed medication.

POPULATION-WIDE INTERVENTION: Many of the interventions in this scenario were modelled by altering the coefficients of the models that were used to estimate the attributes of the synthetic individuals. Specifically, this approach was followed for BMI and SBP. Smoking and fruit and vegetable consumption interventions were modelled by altering the attributes of synthetic individuals after they were estimated in step 2 in figure 2.1 on page 45. Given the existing limitations to measure the direct effect of a structural population-wide intervention, I inflated the uncertainty around the inputs I used for this scenario (table A.1 on page 187).

C.4   SENSITIVITY ANALYSIS RESULTS, PERTAINING CHAPTER 5

Here I present the full output of the three scenarios that were produced as variations of the main scenarios with modified assumptions; namely the '20% treatment threshold universal screening', the 'socioeconomic differential uptake universal screening', and the 'diet only population-wide intervention'. Table C.2, table C.3 on the next page, and table C.4 on the facing page summarise the results.

Table C.2: Estimated cases and deaths prevented or postponed under each scenario, by 2030. Brackets contain the respective interquartile ranges (IQRs).

Scenarios	Cases prevented or postponed by 2030 (IQR)	Deaths prevented or postponed by 2030 (IQR)
20% treatment threshold universal screening	7000 (−2000 to 15 000)	2300 (−1200 to 5600)
Socioeconomic differential uptake universal screening	19 000 (10 000 to 27 000)	700 (−2400 to 4000)
Diet only population-wide intervention	47 000 (37 000 to 56 000)	5600 (2700 to 8700)

Table C.3: Cases prevented or postponed per quintile groups of the Index of Multiple Deprivation, by 2030. The absolute equity slope index for each scenario is also presented. Brackets contain the respective interquartile ranges (IQRs).

Quintile groups of Index of Multiple Deprivation	20% treatment threshold universal screening	Socioeconomic differential uptake universal screening	Diet only population-wide intervention
1 (least deprived)	1400 (−3600 to 6200)	3200 (−1800 to 7800)	8600 (4100 to 13 400)
2	700 (−4800 to 5500)	3900 (−900 to 9000)	9100 (4000 to 13 900)
3	1100 (−4100 to 6700)	4400 (−1500 to 9400)	9400 (4700 to 14 600)
4	1100 (−3500 to 6400)	3400 (−1300 to 8900)	9100 (4200 to 13 700)
5 (most deprived)	2900 (−2800 to 8400)	4300 (−1300 to 9600)	10 400 (5500 to 15 800)
Absolute equity slope index	2000 (−6700 to 10 600)	300 (−6900 to 9000)	2200 (−5300 to 9900)

Table C.4: Relative percentage reduction in cardiovascular disease cases per quintile groups of the Index of Multiple Deprivation, by 2030. The relative equity slope index for each scenario is also presented. Brackets contain the respective interquartile ranges (IQRs).

Quintile groups of Index of Multiple Deprivation	20% treatment threshold universal screening	Socioeconomic differential uptake universal screening	Diet only population-wide intervention
1 (least deprived)	0.6 % (−1.4 % to 2.4 %)	1.2 % (−0.7 % to 3.0 %)	3.3 % (1.6 % to 5.1 %)
2	0.2 % (−1.6 % to 1.9 %)	1.4 % (−0.4 % to 3.1 %)	3.1 % (1.4 % to 4.8 %)
3	0.4 % (−1.4 % to 2.3 %)	1.5 % (−0.5 % to 3.2 %)	3.3 % (1.6 % to 5.0 %)
4	0.4 % (−1.3 % to 2.3 %)	1.2 % (−0.5 % to 3.2 %)	3.3 % (1.5 % to 5.0 %)
5 (most deprived)	1.0 % (−0.9 % to 2.7 %)	1.4 % (−0.4 % to 3.1 %)	3.4 % (1.8 % to 5.3 %)
Absolute equity slope index	0.6 (−2.4 to 3.8)	0.0 (−2.6 to 3.0)	0.4 (−2.2 to 3.0)

C.5   PUBLISHED PEER-REVIEWED PAPERS THAT DIRECTLY STEMMED FROM MY THESIS

RESEARCH ARTICLE

Quantifying the Contribution of Statins to the Decline in Population Mean Cholesterol by Socioeconomic Group in England 1991 - 2012: A Modelling Study

Chris Kypridemos[1‡]*, Piotr Bandosz[1,2‡], Graeme L. Hickey[3‡], Maria Guzman-Castillo[1‡], Kirk Allen[1,4‡], Iain Buchan[5‡], Simon Capewell[1‡], Martin O'Flaherty[1‡]

1 Department of Public Health and Policy, University of Liverpool, Liverpool, United Kingdom, 2 Department of Hypertension and Diabetology, Medical University of Gdansk, Gdansk, Poland, 3 Epidemiology and Population Health Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom, 4 Lancaster Medical School, Lancaster University, Lancaster, United Kingdom, 5 Centre for Health Informatics, Institute of Population Health, University of Manchester, Manchester, United Kingdom

‡ All authors contributed equally to this work.
* c.kypridemos@liverpool.ac.uk

Abstract

Background

Serum total cholesterol is one of the major targets for cardiovascular disease prevention. Statins are effective for cholesterol control in individual patients. At the population level, however, their contribution to total cholesterol decline remains unclear. The aim of this study was to quantify the contribution of statins to the observed fall in population mean cholesterol levels in England over the past two decades, and explore any differences between socioeconomic groups.

Methods and Findings

This is a modelling study based on data from the Health Survey for England. We analysed changes in observed mean total cholesterol levels in the adult England population between 1991-92 (baseline) and 2011-12. We then compared the observed changes with a counterfactual 'no statins' scenario, where the impact of statins on population total cholesterol was estimated and removed. We estimated uncertainty intervals (UI) using Monte Carlo simulation, where confidence intervals (CI) were impractical. In 2011-12, 13.2% (95% CI: 12.5-14.0%) of the English adult population used statins at least once per week, compared with 1991-92 when the proportion was just 0.5% (95% CI: 0.3-1.0%). Between 1991-92 and 2011-12, mean total cholesterol declined from 5.86 mmol/L (95% CI: 5.82-5.90) to 5.17 mmol/L (95% CI: 5.14-5.20). For 2011-12, mean total cholesterol was lower in more deprived groups. In our 'no statins' scenario we predicted a mean total cholesterol of 5.36 mmol/L (95% CI: 5.33-5.40) for 2011-12. Statins were responsible for approximately 33.7% (95% UI: 28.9-38.8%) of the total cholesterol reduction since 1991-92. The statin

contribution to cholesterol reduction was greater among the more deprived groups of women, while showing little socio-economic gradient among men.

Conclusions

Our model suggests that statins explained around a third of the substantial falls in total cholesterol observed in England since 1991. Approximately two thirds of the cholesterol decrease can reasonably be attributed non-pharmacological determinants.

Introduction

Cardiovascular disease (CVD) remains the primary cause of death in the UK and globally [1]. However, UK cardiovascular mortality has been falling consistently since the early 1970s [2]. The two main drivers of this fall have been: reductions in cardiovascular risk factor levels; and improved treatments, both preventive and therapeutic [3].

Serum total cholesterol is one of the main targets for primary and secondary prevention of CVD. In England, the mean total cholesterol of the population has dropped substantially over the past three decades [4]. This fall occurred initially as the result of dietary changes alone [5], but more recently it reflects the interplay between improving diet and increasing statin use [6]. Unlike other cardiovascular risk factors, total cholesterol shows no socioeconomic gradient in young adults and an inverse gradient at older ages, thus more affluent groups appear to have higher total cholesterol levels, especially since 1998 [7].

Despite a plethora of information on the effectiveness of statins at the individual level, especially for secondary prevention, their contribution to the total cholesterol fall in the wider population remains unclear. Farzadfar et al. and Cohen et al. suggest that statins are important in lowering population mean total cholesterol in high income countries including the United States (US) [8,9]. However, it seems that this is neither completely true, nor universal because: 1) large falls in total cholesterol occurred before statins were widely used [10,11]; and 2) the large recent total cholesterol falls observed in Iceland, Sweden, Czech and Finland are principally attributed to improved diets [12–15]. In addition, there are policy concerns over statins and health inequalities. This is because statin prescription is a healthcare based intervention, requiring individual action, which might potentially increase inequalities [16,17].

The debate about statins for primary prevention of CVD has become heated. Last year, the American College of Cardiology (ACC) and the American Heart Association (AHA) updated their recommendations for the treatment of total cholesterol, substantially widening the criteria for statin prescription in otherwise healthy individuals [18]. Now, the UK National Institute for Health and Care Excellence (NICE) has made similar recommendations to drop the ten-year annual risk threshold from 20% to 10%, and almost double the number of eligible adults, from 7 million to 12 million [19]. This has proved very controversial [20,21].

The primary objective of this study was to quantify the contribution of statins to the observed fall in population mean cholesterol levels in England over the past two decades. A secondary objective was to look for any differences in this contribution between socioeconomic groups.

Methods

We analysed changes in observed mean total cholesterol levels in the adult England population between 1991–92 (baseline) and 2011–12. We then compared the observed changes with a hypothetical counterfactual 'no statins' scenario, where the impact of statins on population total

cholesterol was estimated and removed. Therefore, the 'no statins' scenario estimates the hypothetical mean cholesterol of the population, if statins were not available and the population had no benefit from them. Any gap between the observed and the estimated mean total cholesterol would then be attributed to all other possible drivers of population cholesterol levels, principally diet. We stratified our analysis by age-group, sex and, where possible and relevant, by quintiles of the 2010 Index of Multiple Deprivation (QIMD) [22].

Survey data

Specifically, we used anonymised, non-identifiable, participant-level data from the Health Survey for England (HSE) for the two respective periods [23–25]. For the 2011–12 period we aggregated the data of HSE 2011 and HSE 2012, while for 1991–92 this was independently performed by HSE analysts. These cross-sectional surveys provide a representative sample of the non-institutionalised population in England for the respective years. The data files contained anonymised, individualised information for all the participants. We excluded participants younger than 18 years old. For HSE 2011–12 both the weighting and the sampling design were considered in the estimation of all the point estimates and their standard errors. In particular, the weighting adjusts both for selection and non-response bias. The sample for HSE 1991–92 was un-weighted, therefore, only the sampling design was taken into account. Further details about HSE can be found elsewhere [26–28].

Socioeconomic stratification

There were no common socioeconomic indicators between the two samples; QIMD was therefore used for the 2011–12 sample and social class based on occupation (I—V) was used for the 1991–92 sample.

QIMD is a measure of relative area deprivation based on the 2010 version of the Index of Multiple Deprivation [22]. According to this system, all Lower Super Output Areas in England (LSOA) (average population of 1,500) are ranked in order of increasing deprivation, based on seven domains of deprivation: income; employment; health deprivation and disability; education, skills and training; barriers to housing and services; crime and disorder, and living environment. For the ranking, individual level information about the habitats of these areas is used from multiple sources. Then, the QIMD is formed from the quintiles of the above index, one through five, where quintile one is considered the 'most affluent' and quintile five the 'most deprived'. The HSE team provided the QIMD of each participant for HSE 2011–12 based on their postcode of residence, which is a sub-division of LSOAs. We opted to use the QIMD instead of other available socioeconomic classification systems mainly for three reasons. First, the QIMD was the only socioeconomic indicator that had no missing cases in our data, second, for our results to be comparable with other studies that used QIMD and third, because QIMD is extensively used by local public health departments, Office of National Statistics and researchers in England.

The HSE 1991–92 social class classification was based on the 1990 version of the Standard Occupational Classification (SOC90) [29] and the self-reported occupation of the participants. Social class was provided as a variable in the data, by the HSE team. We aggregated full time students, armed forces personnel, those who never worked, and those whose occupation was not fully described in one category ('Other'). In our analysis, we avoided any direct comparisons between the two socioeconomic classification systems.

Total cholesterol measurement

Total cholesterol is reported in millimoles per litre (mmol/L). To convert it to milligrams per decilitre (mg/dL) please multiply the reported cholesterol values by 38.6. In 2011–12 a sub-

sample of the total HSE sample was eligible and consented to provide non-fasting blood samples for the measurement of total cholesterol in serum. For HSE 1991–92, participants aged 18 and over were asked to provide a blood sample for the same purpose. Since April 2010 the equipment that was used for the measurement of total cholesterol for HSE was replaced. The effect of this change was that measured concentrations of total cholesterol from this date onwards were on average 0.1mmol/L higher. We adjusted for this difference in our analyses by subtracting 0.1mmol/L from the respective total cholesterol measurements. A more detailed description of the total cholesterol measurement process can be found elsewhere (pages 32–36 in [26], and pages 31–35 in [27]).

Estimating statin utilisation

In England, individuals may have access to statins using two available routes. Statins can either be prescribed to them by a doctor (or a non-medical prescriber), or they can be bought over the counter (OTC) from a pharmacy with or without prior expert advice. HSE assessed both routes. In 2011–12, during a nurse interview, the participants were asked to report the medication that had been prescribed to them by a doctor or by a non-medical prescriber. Specifically for statins, they were also asked whether they bought OTC. Finally, those that had been prescribed a statin or bought it OTC were asked if they had used it during the past seven days. We only considered the participants that answered positively in the last question as statin users. For HSE 1991–92 the participants were asked similar questions during the nurse interview. However, statins were included in the wider category of lipid-lowering medication and were not prescribed for primary prevention [30,31]. Since the uptake of this category as a whole was very low, we assumed that statins had a negligible effect on total cholesterol at population level; thus, we ignored it completely (please see S1 Text for further justification of this assumption).

Statistical analysis

The analysis was performed in R statistical software (v3.1.0) [32] including the R package "survey" [33]. An approximate 95% confidence interval (CI) for proportions (e.g. statin uptake) was calculated from the survey data using the incomplete beta function method, with an effective sample size based on the estimated variance of the proportion [34]. Missing cases were excluded from our analysis (please refer to Table 1).

To test the statistical significance of socioeconomic trends in total cholesterol, against the null hypothesis of 'no trend', we fitted a generalised linear model, with inverse-probability weighting and design-based standard errors. Specifically, we treated total cholesterol measurements as the dependent variable and the QIMD (or social class) as the independent one. We considered QIMD and social class as numeric variables for this (e.g. QIMD 1 through 5 represented the 5 quintiles and social class 1 through 7 represented the social classes I, II, IIIN, IIIM, IV, V and 'Other' respectively). Therefore, the β coefficient (slope) of the QIMD (or social class) and its standard error was a measure of the socioeconomic gradient. When β was not statistically significant we assumed no socioeconomic gradient. When β was statistically significant, its sign revealed the direction of the gradient (e.g. a negative sign means that mean total cholesterol is lower among the more deprived groups) and its absolute value measured the magnitude of the gradient.

A similar approach was followed to explore socioeconomic trends in statin utilisation. Since this time the dependent variable was a binary one, we used a binomial model.

Estimating the effect of statins. The average effect of each specific statin and strength on an individual's total cholesterol is known from the literature [35–38]. However, the exact type of statin, and strength, had not been recorded for the participants in HSE 2011–12. To

Table 1. Samples baseline characteristics. Values are numbers (percentages).

Characteristics	Number of participants interviewed by a nurse				Number of participants with a valid total cholesterol result			
	1991–92 (n = 7043)		2011–12 (n = 10965)		1991–92 (n = 4995)		2011–12 (n = 7772)	
	Men	Women	Men	Women	Men	Women	Men	Women
Age (years)								
18–34	999 (14.2)	1165 (16.5)	877 (8.0)	1350 (12.3)	733 (14.7)	730 (14.6)	604 (7.8)	797 (10.3)
35–54	1148 (16.3)	1240 (17.6)	1632 (14.9)	2194 (20.0)	886 (17.7)	921 (18.4)	1216 (15.6)	1633 (21.0)
55+	1101 (15.6)	1390 (19.7)	2254 (19.7)	2658 (24.2)	806 (16.1)	919 (18.4)	1611 (20.7)	1911 (24.6)
QIMD								
1 (most affluent)	-	-	1058 (9.6)	1389 (12.7)	-	-	785 (10.1)	995 (12.8)
2	-	-	1057 (9.6)	1364 (12.4)	-	-	791 (10.2)	997 (12.8)
3	-	-	1017 (9.3)	1278 (11.7)	-	-	732 (9.4)	892 (11.5)
4	-	-	865 (7.9)	1133 (10.3)	-	-	606 (7.8)	781 (10.0)
5 (most deprived)	-	-	766 (7.0)	1038 (9.5)	-	-	517 (6.7)	676 (8.7)
Social class								
I Professional	235 (3.3)	53 (0.8)	-	-	174 (3.5)	41 (0.8)	-	-
II Managerial technical	908 (12.9)	856 (12.2)	-	-	688 (13.8)	610 (12.2)	-	-
IIIN Skilled non-manual	320 (4.5)	1304 (18.5)	-	-	238 (4.8)	909 (18.2)	-	-
IIIM Skilled manual	1085 (15.4)	388 (5.5)	-	-	816 (16.3)	251 (5.0)	-	-
IV Semi-skilled manual	460 (6.5)	693 (9.8)	-	-	343 (6.9)	464 (9.3)	-	-
V Unskilled manual	157 (2.2)	363 (5.2)	-	-	112 (2.2)	225 (4.5)	-	-
Other	83 (1.2)	138 (2.0)	-	-	54 (1.1)	70 (1.4)	-	-

The difference between the number of participants that had a nurse interview and those who had a valid total cholesterol result indicates the missing cases. QIMD denotes quintiles of index of multiple deprivation (1 = most affluent, 5 = most deprived).

doi:10.1371/journal.pone.0123112.t001

overcome this limitation we used the exact amount of statins (by proprietary name and strength) that were both prescribed and dispensed in England for 2011 and 2012, available from the Health and Social Care Information Centre [39,40]. We then estimated a weighted mean of the proportional decrease of total cholesterol attributable to statins overall (Eq 1).

$$E_w = \frac{\sum_i \sum_j (M_{ij} * E_{ij})}{\sum_i \sum_j (M_{ij})} \tag{1}$$

Eq 1. Formula for the estimation of the proportional decrease in mean total cholesterol attributable to overall statins use.

Where:

E_w is the proportional decrease in mean total cholesterol attributable to statins, among statin users

E_{ij} is the proportional decrease in mean total cholesterol attributable to a specific statin i of a specific strength j (e.g. Simvastatin 20mg)

M_{ij} is the number of units of a specific statin i and strength j that have been prescribed and dispensed. For liquid forms 5ml were considered as one unit, otherwise one tablet was considered as a unit

For the estimation of E_{ij} data from several meta-analysis were used as follows: We obtained the mean and standard error (calculated directly from the 95% CI assuming approximate

normality) of the proportional reduction in serum low-density lipoprotein (LDL) from the meta-analysis of Law et al. [35]. The proportional reduction was derived from the absolute reduction, standardised to usual serum LDL of 4.8 mmol/L before treatment, and it was independent of the pre-treatment LDL. This allowed us to use a weighted mean approach on proportions. We then converted the LDL reduction into total cholesterol reduction using data from other studies, [36–38] assuming a linear relation between total cholesterol and LDL reduction. For strengths not included in the above meta-analysis (e.g. Atorvastatin 30mg), we used a linear regression model to estimate their effect, based on the effect of known strengths. Specifically, we treated the total cholesterol reduction as the dependent variable and the natural logarithm of strength as the independent one. We weighted the model against the inverse variance of the cholesterol reduction. The effectiveness of solid and liquid forms was considered equal. Similarly, the effectiveness of the combined forms of simvastatin with ezetimibe was considered equal to the effectiveness of same strength simvastatin (S1 Table). The standard error of E_w was estimated using the Cochran's definition for the standard error of the weighted mean [41,42].

For the 'no statins' scenario, we calculated the predicted total cholesterol for each statin user, with the effect of statin removed using the formula below (Eq 2).

$$TC_{pred} = \frac{TC_{obs}}{1 - E_w} \qquad (2)$$

Eq 2. Formula for the calculation of predicted total cholesterol with the effect of statins removed.

Where:

TC_{pred} is the predicted total cholesterol of the statin user with the statin effect removed

TC_{obs} is the observed total cholesterol of the statin user

E_w is the proportional decrease in mean total cholesterol attributable to statins, derived from Eq 1.

We used Monte Carlo simulation to incorporate the uncertainty from the sampling distribution of E_w. For each statin user we drew 1000 values from a normal distribution with mean E_w and standard deviation as per the estimated standard error (described above). We then averaged over the TC_{pred} predictions and considered this mean value as the predicted total cholesterol of each statin user, with the statin effect removed.

Quantifying the contribution of statins on population's mean total cholesterol reduction. To quantify and compare the contribution of statins against the contribution of all other total cholesterol lowering interventions in the population, we first plotted the mean total cholesterol for 1991–92, 2011–12 and the 'no statins scenario' by age for each sex. We considered the area enclosed by the respective curves for 1991–92 and 2011–12 as representing the full observed cholesterol reduction (area A). Therefore, the area enclosed by the 2011–12 and the 'no statin' scenario represents the reduction of cholesterol attributable to statins (area B). Thus, the fraction (area B) / (area A) expresses the contribution of statins to the observed decline of mean total cholesterol. For the estimation of areas A and B we used natural spline interpolation as implemented in the R package "MESS" [43].

To estimate the uncertainty intervals (UI) around the estimated contribution of statins, we modified the previous method to allow for a Monte Carlo simulation approach. Specifically, for each age in the population, we drew 10000 values from the conditional sampling distribution, which we approximated by a normal distribution with age-specific estimate mean and standard error. These are then averaged across the age range to yield a point estimate, and 2.5%

and 97.5% percentiles were used to define the 95% UI. Due to small representation of ages above 89 in our sample, we aggregated participants older than 89 years with those aged 89.

Finally, we repeated the analysis separately for each QIMD under the assumption that total cholesterol had no socioeconomic gradient in 1991–92. We further limited the analysis in participants younger than 76 years because of the small number of older participants in our sample, when stratified by QIMD. To test the statistical significance of any observed socioeconomic trend we used the two-tailed Cochran-Armitage trend test.

Sensitivity analysis. For the estimation of E_w several assumptions were involved that do not necessarily reflect on its estimated standard error. We repeated our analysis after we multiplied the standard error of E_w by a factor of 10 in order to test the robustness of our results with a higher than measured uncertainty scenario.

Ethical approval

Ethical approval for the 2011 and 2012 surveys was obtained from the Oxford A Research Ethics Committee (reference numbers 10/H0604/56) by the Health Survey for England team. For 1991 and 1992 surveys ethical approval had been granted by the Local Research Ethics Councils in England. Anonymised, non-identifiable data of HSE are available to academics and public sector staff through the UK Data Archive (www.data-archive.ac.uk) for secondary analysis, without requiring further approval.

Results

The baseline characteristics of the 1991–92 and 2011–12 samples are summarised in Table 1, while mean total cholesterol values by age group and sex are presented in Table 2 (1991–92) and Table 3 (2011–12). Overall, the prevalence of statin use in England, including OTC statin

Table 2. Observed mean total cholesterol (mmol/L) overall, and by age group, sex and social class in England, 1991–92.

Social class	18–34 (years)		35–54		55+		Overall
	Men	Women	Men	Women	Men	Women	
I Professional	5.52 (5.20 to 5.83)	5.10 (4.70 to 5.50)	5.95 (5.71 to 6.19)	5.64 (5.26 to 6.03)	5.99 (5.66 to 6.31)	6.62 (6.12 to 7.12)	5.64 (5.48 to 5.81)
II Managerial technical	5.25 (5.06 to 5.44)	5.05 (4.93 to 5.17)	6.01 (5.89 to 6.13)	5.57 (5.46 to 5.69)	6.24 (6.10 to 6.39)	6.79 (6.62 to 6.97)	5.69 (5.58 to 5.82)
IIIN Skilled non-manual	5.24 (5.06 to 5.43)	5.02 (4.92 to 5.12)	6.15 (5.88 to 6.41)	5.71 (5.58 to 5.83)	6.08 (5.81 to 6.36)	6.80 (6.66 to 6.94)	5.64 (5.49 to 5.79)
IIIM Skilled manual	5.16 (5.04 to 5.27)	5.05 (4.81 to 5.29)	5.93 (5.78 to 6.07)	5.97 (5.70 to 6.24)	6.06 (5.95 to 6.18)	6.83 (6.61 to 7.05)	5.72 (5.61 to 5.84)
IV Semi-skilled manual	5.16 (4.95 to 5.37)	5.12 (4.96 to 5.27)	5.89 (5.68 to 6.11)	5.70 (5.53 to 5.87)	6.00 (5.82 to 6.19)	6.95 (6.76 to 7.14)	5.70 (5.55 to 5.85)
V Unskilled manual	5.25 (4.82 to 5.68)	5.15 (4.84 to 5.45)	6.07 (5.67 to 6.47)	6.00 (5.77 to 6.22)	6.04 (5.63 to 6.45)	6.97 (6.54 to 7.41)	6.00 (5.79 to 6.21)
Other	4.70 (4.39 to 5.01)	5.14 (4.82 to 5.46)	5.82 (5.03 to 6.61)	5.03 (4.48 to 5.57)	6.37 (5.62 to 7.13)	6.70 (6.14 to 7.26)	5.27 (5.06 to 5.49)
All	5.20 (5.12 to 5.27)	5.06 (5.00 to 5.13)	5.97 (5.90 to 6.05)	5.70 (5.64 to 5.77)	6.10 (6.03 to 6.18)	6.84 (6.76 to 6.93)	
Slope of the trend	-0.07 (-0.13 to -0.01)	0.02 (-0.02 to 0.07)	-0.02 (-0.07 to 0.03)	0.05 (0.00 to 0.10)	-0.04 (-0.10 to 0.02)	0.04 (-0.04 to 0.11)	0.00 (-0.02 to 0.02)
***P* for trend**	**0.01**	0.27	0.47	**0.03**	0.19	0.32	0.96

Socioeconomic trends are also presented. Brackets contain 95% confidence intervals. The 'Overall' column is adjusted for age and sex.

doi:10.1371/journal.pone.0123112.t002

Table 3. Observed mean total cholesterol (mmol/L) overall, and by age group, sex and quintiles of index of multiple deprivation (QIMD) (1 = most affluent, 5 = most deprived) in England, 2011–12.

QIMD	18–34 (years)		35–54		55+		Overall
	Men	Women	Men	Women	Men	Women	
1 (most affluent)	4.80 (4.60 to 5.00)	4.76 (4.60–4.92)	5.53 (5.42 to 5.64)	5.24 (5.13 to 5.36)	5.12 (5.01 to 5.23)	5.77 (5.67 to 5.87)	5.19 (5.09 to 5.29)
2	4.71 (4.56 to 4.86)	4.46 (4.31 to 4.61)	5.47 (5.33 to 5.61)	5.19 (5.08 to 5.31)	5.07 (4.95 to 5.19)	5.72 (5.61 to 5.82)	5.09 (4.99 to 5.20)
3	4.63 (4.41 to 4.86)	4.70 (4.53 to 4.87)	5.64 (5.50 to 5.79)	5.26 (5.15 to 5.38)	5.05 (4.91 to 5.18)	5.67 (5.54 to 5.80)	5.10 (4.99 to 5.22)
4	4.84 (4.65 to 5.02)	4.61 (4.46 to 4.77)	5.46 (5.30 to 5.62)	5.35 (5.20 to 5.49)	4.95 (4.80 to 5.11)	5.55 (5.40 to 5.70)	5.05 (4.94 to 5.17)
5 (most deprived)	4.79 (4.57 to 5.01)	4.59 (4.44 to 4.74)	5.40 (5.24 to 5.57)	5.31 (5.17 to 5.45)	4.74 (4.55 to 4.92)	5.34 (5.15 to 5.54)	4.93 (4.82 to 5.05)
All	4.75 (4.66 to 4.84)	4.62 (4.55 to 4.69)	5.50 (5.44 to 5.57)	5.26 (5.21 to 5.32)	5.02 (4.96 to 5.08)	5.64 (5.58 to 5.70)	
Slope of the trend	0.02 (-0.05 to 0.08)	-0.01 (-0.06 to 0.04)	-0.03 (-0.07 to 0.02)	0.03 (-0.01 to 0.07)	-0.08 (-0.12 to -0.03)	-0.10 (-0.14 to -0.05)	-0.03 (-0.05 to -0.01)
P for trend	0.60	0.67	0.26	0.16	**<0.001**	**<0.001**	**0.002**

Socioeconomic trends are also presented. The 'Overall' column is adjusted for age and sex. Brackets contain 95% confidence intervals.

doi:10.1371/journal.pone.0123112.t003

users was 13.2% (95% CI: 12.5% to 14.0%) in 2011–12. Another 0.8% (95% CI: 0.6% to 1.0%) of the population were prescribed or bought OTC statins; however, they did not use them for at least a week before the nurse interview.

For 1991–92, statin use was not specifically recorded in the survey; however, the prevalence of all lipid lowering medications, including statins, was 0.5% (95% CI: 0.3% to 1.0%). Table 4 summarises the prevalence of statin use in England for 2011–12 by age group, sex and QIMD. There was a statistically significant socioeconomic gradient in ages above 35 years for both sexes, where the use of statins increased with deprivation.

In 2011–12, some 13.1% (95% CI: 12.4 to 14.0%) of study population used statins prescribed to them (not including OTC users), over the seven days before the survey interview. We estimated the expected number of units (e.g. tablets or 5ml doses of liquid statins) that were consumed in England for the same period, assuming that they stayed on statins for the whole year and that institutionalised population shares the same consumption attitudes, to be

Table 4. Prevalence of statin use in England 2011–12 by age, sex and quintiles of index of multiple deprivation (QIMD) (1 = most affluent, 5 = most deprived).

QIMD	18–34 (years)		35–54		55+		Overall
	Men	Women	Men	Women	Men	Women	
1 (most affluent)	0% (0–2%)	-	5% (3–8%)	2% (1–3%)	36% (32–41%)	20% (16–23%)	19% (16–23%)
2	-	-	7% (4–11%)	3% (2–5%)	38% (34–43%)	24% (20–27%)	22% (18–26%)
3	0% (0–2%)	0% (0–2%)	7% (5–11%)	2% (1–4%)	32% (28–37%)	29% (25–33%)	20% (16–24%)
4	1% (0–5%)	-	8% (5–12%)	4% (2–6%)	39% (34–44%)	29% (25–34%)	20% (17–24%)
5 (most deprived)	-	1% (0–3%)	9% (6–13%)	8% (5–11%)	47% (40–54%)	34% (29–40%)	21% (17–25%)
All	0% (0–1%)	0% (0–1%)	7% (6–9%)	4% (3–4%)	38% (36–40%)	26% (25–28%)	
P for trend	-	-	**0.03**	**< 0.001**	**0.04**	**<0.001**	**<0.001**

The 'Overall' column is adjusted for age and sex. Brackets contain 95% confidence intervals.

doi:10.1371/journal.pone.0123112.t004

approximately 4.00 billion. This showed reassuringly close agreement with the observed unit consumption of almost 4.07 billion [39,40], being just 1.5% lower.

The mean total cholesterol of adult non-institutionalised population in England decreased from 5.86 mmol/L (95% CI: 5.82 to 5.90) in 1991–92 to 5.17 mmol/L (95% CI: 5.14 to 5.20) in 2011–12. The decrease was observed in all age groups and it was steeper for ages over 55 for women and 35 for men (Fig 1). The inverse socioeconomic gradient observed since 1998 [7] persisted overall and in the subgroup of those aged over 55 years. No gradient was observed for other age groups (Table 3). On the contrary, we did not observe any socioeconomic gradient in 1991–92 with social class as a socioeconomic indicator when adjusted for age and sex (Table 2). The trend remained non-significant even when we placed the 'Other' social class group before all other groups.

'No statins' scenario

We estimated the total effect of statins on total cholesterol reduction using Eq 1 as E_w = 25.7% (95% CI: 23.3% to 28.0%). The mean predicted total cholesterol TC_{pred} of the population was calculated to be 5.36 mmol/L (95% CI: 5.33 to 5.40).

Fig 2 depicts the predicted mean total cholesterol of the population without the effect of statins, against the observed mean total cholesterol in 1991–92 and 2011–12, by age and sex. When the effect of statins was removed, the inverse socioeconomic gradient of cholesterol in the overall population disappeared (slope -0.01, 95% CI: -0.03 to 0.01, P = 0.45). Subgroup analysis revealed that for men over 55 the slope was reduced to -0.05 (95% CI: -0.10 to -0.01, $P = 0.03$) and for women over 55 the gradient was essentially zero (slope -0.04, 95% CI: -0.08 to 0.01, P = 0.09). In addition, a socioeconomic trend appeared for women between 35 and 54 years with a slope of 0.05 (95% CI: 0.01 to 0.10, P = 0.01). We saw no other statistically significant gradient, for the remaining age groups (S2 Table).

Finally, statins were estimated as responsible for approximately 33.7% (95% UI: 28.9% to 38.8%) of the total cholesterol reduction since 1991–92. When stratified by sex statins contribution was 40.1% (95% UI: 33.6% to 47.7%) in men and 28.6% (95% UI: 22.3% to 35.0%) in women. Table 5 summarises the contribution of statins for each socioeconomic group, by age group and sex. The negative values in the UI, implying that statins could have increased cholesterol to some, are an artefact of the Monte Carlo simulation due to wide mean cholesterol CI overlapping in some ages. Statins' contribution was consistently higher among men, consistent with the observed higher utilisation.

Sensitivity analysis

The mean predicted total cholesterol (TC_{pred}) of the population, using the inflated standard error of E_w, was calculated to be 5.39 mmol/L (95% CI: 5.35 to 5.42). This is less than a 0.03 mmol/L difference from the main analysis. For the subgroup of deprived men older than 55, with the highest statin utilisation, the TC_{pred} from the sensitivity analysis was 0.09 mmol/L higher than the one from the main analysis. Similarly, the contribution of statins to the observed cholesterol decline for the whole population was estimated to be 33.9% (95% UI: 28.8 to 38.7%), a 0.2% difference from the main analysis result. A similar pattern of minimal changes was observed for the remaining results.

Discussion

This is the first study we know of to quantify the contribution of statins to the observed decrease of total cholesterol in England's population by socioeconomic group. Our results strongly suggest that the statins were not the main driver of total cholesterol reduction since 1991–92.

Mean total cholesterol by age group and sex
England 1991-92 and 2011-12
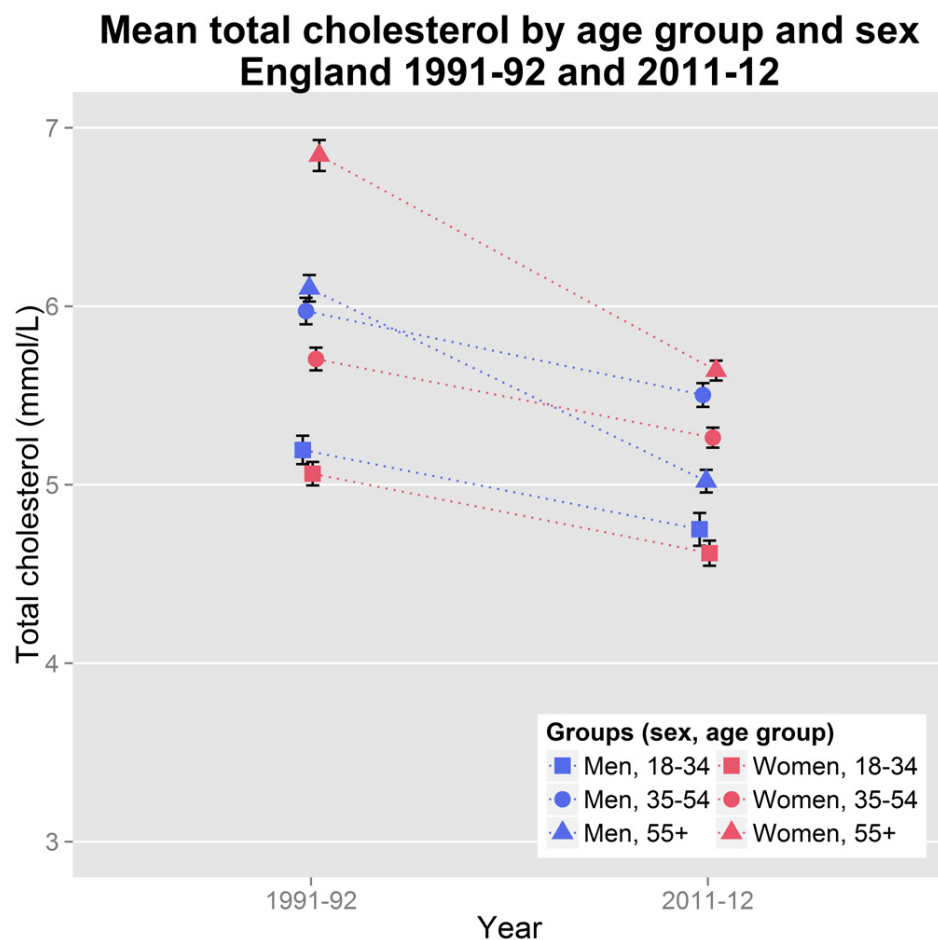
Fig 1. Mean serum total cholesterol (mmol/L) observed decline in England from 1991–92 to 2011–12 in men and women by age group. The error bars depict 95% confidence interval of the means. The vertical axis starts at 3 mmol/L to improve readability. The dotted lines are visual aids and do not reflect linear fits.

doi:10.1371/journal.pone.0123112.g001

In fact, only around one third of the overall reduction might be attributed to statins, and that was mainly in patients aged over 55 years. Statins were more widely used in deprived than affluent areas. They appeared to help reduce socioeconomic inequalities in total cholesterol among women, but not among men.

Obseved and predicted mean total cholesterol by age, in men and women
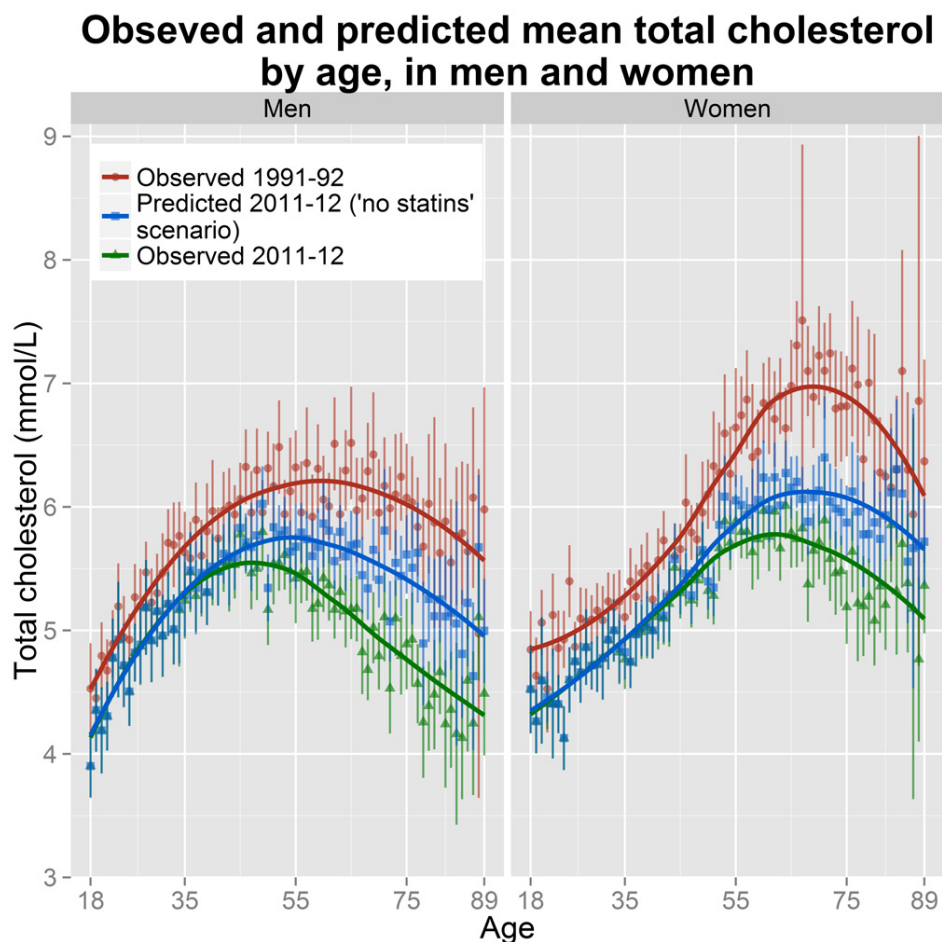
Fig 2. Mean serum total cholesterol by age, in men and women, in England (observed and predicted values). The points depict the mean total cholesterol and the vertical lines 95% confidence intervals (CI). The curves were derived from weighted local regressions and are used to enhance readability. Due to small sample sizes we aggregated participants aged 89 with those older than 89 years. To improve readability the axes are not numbered from 0.

doi:10.1371/journal.pone.0123112.g002

Statins utilisation

In our study, statins' utilisation was higher in more deprived areas for men and women aged over 35 years. This socio-economic pattern may partly reflect the higher prevalence of CVD in more deprived areas [44] and the incentivised use of the QRISK score for cardiovascular risk

Table 5. Estimated proportional contribution of statins to total cholesterol reduction since 1991–92 for each quintile of index of multiple deprivation (QIMD), by age group and sex.

QIMD	35–54 (years)		55–75		18–75	
	Men	Women	Men	Women	Men	Women
1 (most affluent)	14.0% (-19.2 to 41.9%)	4.2% (-24.4 to 28.3%)	50.6% (36.2 to 64.6%)	24.4% (11.4 to 36.6%)	33.5% (15.6 to 49.9%)	15.9% (1.9 to 28.9%)
2	13.0% (-28.2 to 45.9%)	5.9% (-23.9 to 30.7%)	59.7% (43.5 to 75.7%)	23.9% (10.2 to 36.8%)	36.0% (19.5 to 51.2%)	14.3% (2.0 to 25.5%)
3	17.9% (-49.7 to 78.2%)	3.6% (-33.9 to 33.3%)	37.8% (21.3 to 52.7%)	36.0% (21.2 to 50.1%)	26.9% (7.3 to 44.6%)	23.5% (7.5 to 37.5%)
4	29.0% (-7.2 to 58.9%)	19.3% (-37.9 to 64.4)	45.0% (27.7 to 60.6%)	36.3% (21.3 to 50.5%)	34.4% (19.0 to 48.9%)	24.8% (9.2 to 38.9%)
5 (most deprived)	31.1% (-9.1 to 63.8%)	37.1% (-19.0 to 79.9%)	43.2% (28.6 to 57.1%)	32.6% (16.0 to 47.9%)	33.8% (19.7 to 46.1%)	33.4% (18.3 to 47.5%)
All	22.2% (4.8 to 39.8%)	11.9% (-4 to 26.1%)	48.0% (40.1 to 56.1%)	40.0% (23.3 to 54.9%)	33.2% (25.8 to 40.6%)	21.3% (14.8 to 28.0%)
P for trend	0.24	**0.03**	0.41	0.17	0.99	**0.02**

Age group 18–34 was omitted as statins' contribution was practically zero. Analysis was restricted to ages younger than 76 due to low number of older participants. Brackets contain 95% uncertainty intervals estimated by Monte Carlo.

doi:10.1371/journal.pone.0123112.t005

stratification in clinics, which includes area deprivation as a risk factor [45,46]. Our findings are consistent with earlier studies that used different methodologies. Ashworth et al. and Wu et al. also found that statin prescription was higher in more deprived areas in the UK [47,48]. This success in tackling inequalities might be attributed to the National Health Service (NHS), since evidence from Australia, Sweden, Denmark and the US [49–52] suggest that statin prescription in these countries has a socioeconomic gradient, with a less than expected utilisation among the more disadvantaged, and potentially increases health inequalities.

Statins contribution to cholesterol decline

The second interesting finding is the contribution of statins to the observed decline of total cholesterol since 1991–92. We found that statins are not the main driver of the cholesterol decline in England, echoing studies from Iceland, Sweden, Finland and the Czech Republic [12–15]. We estimated that only about a third of the observed total cholesterol decline could be attributed to statins. This contribution was slightly higher than the aforementioned studies, perhaps reflecting a more recent time period with correspondingly higher statin use in England 2011–12, and possible nuanced differences in methodologies. While the cholesterol decrease was observed in all age groups since 1991, statins mostly contributed to the fall in people older than 55 years.

The observed inverse socioeconomic gradient in total cholesterol levels might be partly attributed to statins. In the 'no statins' scenario the gradient disappeared completely when all ages were considered. However, the statin contribution varied across different genders and socioeconomic groups. Statin utilisation was higher in the most deprived groups, but inequitable by gender, reaching barely one third in women (34%) but almost half (47%) of deprived men in the 55+ age group. This difference can only partly be explained by the higher CVD prevalence among men. By contrast, the statin contribution to cholesterol lowering was rather stable across socio-economic groups in men (some 33%), but rose from 16% to 33% in women. This suggests that the component of all other cholesterol reduction drivers had a higher impact among the most deprived men, while their effect among women of all socioeconomic background was more or less equal. This demands further research.

Public health implications

Overall, our research supports the principle of statins being the second best option for primary prevention. Non-statin interventions account for two thirds of the total cholesterol reduction observed since 1991–92, which can be mostly attributed to dietary changes because physical activity levels have not increased substantially over this period [30,53] and the contribution of other factors affecting lipids is small and remained more or less stable. Indeed, United Nations Food and Agriculture Organization data indicate that the animal fat supply per capita in the UK has fallen by almost 25% since 1991 [54]. This echoes Rose's original assertion that the greatest public health impact will be achieved through population-wide reductions in CVD risk than through interventions targeting high-risk individuals [55].

Furthermore, the recent proposed widening of criteria for statin prescription in primary prevention by the ACC/AHA [18] and NICE [19] has been questioned on grounds of effectiveness, cost-effectiveness, acceptability and safety [21]. These measures may prove to be less effective than anticipated because of cumulative attrition factors. Approximately half of the UK patients that are commenced on lipid lowering medication for primary prevention are ineligible according to the respective guidelines, while many eligible patients remain untreated [48]. Moreover, over half the patients commenced on statins for primary prevention have discontinued them within 1–2 years [56–59]. In addition to medicalising otherwise healthy individuals, some patients may also be tempted to adopt more unhealthy diets because of the false 'reassurance' that statins will compensate for the unhealthy behaviours [60]. Along with the increased resource requirements, an additional opportunity cost comes from undermining the primary driver of cholesterol decline—nutritional improvements at individual and national policy levels [61].

Regarding inequalities in health and inequities in care: our research suggests that English statin prescribing might be equitable. This represents a success for the socialised medicine provided by the NHS England. In contrast, statin-based cholesterol reduction was not equitable among men, being similar in the more affluent and more deprived groups. These results are intriguing, because healthcare-based interventions generally increase the inequality gap [16,17].

Strengths and limitations

This study was grounded on the best available evidence to explore the research question. We integrated all the available data from HSE, a cross-sectional survey of very high quality, the Prescription Cost Analysis report, an accurate and precise report about prescriptions in England, and published meta-analyses on the effect of statins. The modelling approach allowed for the best use of all the available information. In fact, despite the assumptions regarding the effects of statins our results were robust to the sensitivity analysis. Any biases and errors were diluted because they only applied to the about 13% of the sample who were statin users.

However, our study has several limitations. First, it is based on self-reported statin prescription and adherence, and does not account for statin indications; however, consistent data from prescription cost analysis reports for 2011–12 [39,40] suggest that our estimated prevalence of statin-use is fairly accurate. Second, unlike HSE 2011–12, HSE 1991–92 was not weighted to adjust for non-response bias. Furthermore, no other HSE has recorded statin use separately from other lipid-lowering medication; this renders an interim point analysis between 1991 and 2011 practically impossible.

Third, there were no common or directly compatible socioeconomic indicators between the two surveys to allow for more accurate comparisons. Our assumption that there was no socioeconomic gradient of mean total cholesterol in 1991–92 is supported by our finding of no such gradient by social class in HSE 1991–92. This is consistent with Scholes et al. who also showed no socioeconomic gradient in 1994 using QIMD as socioeconomic indicator [7]. The Whitehall

II cohort also showed no socioeconomic gradient for total cholesterol in 1985–88 [62]. Neither did our analysis consider other inequalities, for instance, ethnic minorities or people with mental health or illiteracy problems [47,63,64].

Fourth, the estimate of the statin effect E_w was derived mostly from short-term trials lasting less than one year. However, Edward et al. have shown that the statins effect remains fairly stable in trials lasting more than one year (Additional file 5 in [37]). In addition, the estimation of E_w assumes that the differences between each trial population and our study sub-population of statin users were the same for each statin.

Fifth, this analysis cannot fully control for other factors that interfere with lipid profiles and their prevalence in the population changed substantially over the last two decades. BMI and diabetes mellitus are possibly the most important of them.

Finally, we used the statins effects reported in clinical trials, acknowledging that this might overestimate the real world efficacy of these drugs (mostly because of selection bias in the trials and reduced compliance in the population). However, this result in an overestimation of the contribution of statins, and thus its real contribution might have been even less than one third.

Conclusions

Our research suggests that statins contributed about one third of the observed total cholesterol decline in England since 1991–92, and that their impact on reducing socioeconomic inequalities in total cholesterol was generally positive. However, the proposed wider indications for statins in primary prevention remains contested.

Further research is now needed to quantify the potential contribution of primary prevention statins to the 'hard' outcomes of cardiovascular morbidity and mortality in the UK. There is sufficient current evidence, however, to justify reconsidering the priorities of different interventions for the primary prevention of CVD.

Supporting Information

S1 Table. Statins effects and weights used for the estimation of the weighted mean E_w.
(DOCX)

S2 Table. Predicted mean total cholesterol (mmol/L) with the statins effect removed. Overall, and by age group, sex and quintiles of index of multiple deprivation (QIMD) (1 = most affluent, 5 = most deprived) in England, 2011–12. Socioeconomic trends are also presented. Brackets contain 95% confidence intervals.
(DOCX)

S1 Text. Supporting the assumption of no statin effect in 1991–1992.
(DOCX)

Author Contributions

Conceived and designed the experiments: PB MOF SC CK. Performed the experiments: CK. Analyzed the data: CK. Contributed reagents/materials/analysis tools: GLH KA MGC IB. Wrote the paper: CK PB GLH MGC KA IB SC MOF.

References

1. Alwan A, Armstrong T, Bettcher D, Branca F, Chisholm D, Ezzati M, et al. (2011) Global status report on noncommunicable diseases 2010. World Health Organization. Available: http://whqlibdoc.who.int/publications/2011/9789240686458_eng.pdf.

2. Townsend N, Wickramasinghe K, Bhatnagar P, Smolina K, Nichols M, Leal J, et al. (2012) Coronary heart disease statistics 2012 edition. British Heart Foundation. Available: http://www.bhf.org.uk/publications/view-publication.aspx?ps=1002097.

3. Unal B, Critchley JA, Capewell S (2004) Explaining the decline in coronary heart disease mortality in England and Wales between 1981 and 2000. Circulation 109: 1101–1107. doi: 10.1161/01.CIR.0000118498.35499.B2 PMID: 14993137

4. The Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (n.d.) Global Burden of Metabolic Risk Factors of Chronic Diseases. Available: http://www1.imperial.ac.uk/publichealth/departments/ebs/projects/eresh/majidezzati/healthmetrics/metabolicriskfactors/. Accessed 6 October 2014.

5. Morton GM, Lee SM, Buss DH, Lawrance P (1995) Intakes and major dietary sources of cholesterol and phytosterols in the British diet. J Hum Nutr Diet 8: 429–440. doi: 10.1111/j.1365-277X.1995.tb00338.x

6. Department of Health (n.d.) Pharmacies and prescriptions: statistics. Available: http://webarchive.nationalarchives.gov.uk/20120503222906/http://www.dh.gov.uk/en/PublicationsAndStatistics/Statistics/StatisticalWorkAreas/StatisticalHealthCare/DH_4086488. Accessed 15 June 2014.

7. Scholes S, Bajekal M, Love H, Hawkins N, Raine R, O'Flaherty M, et al. (2012) Persistent socioeconomic inequalities in cardiovascular risk factors in England over 1994–2008: A time-trend analysis of repeated cross-sectional data. BMC Public Health 12: 129. doi: 10.1186/1471-2458-12-129 PMID: 22333887

8. Farzadfar F, Finucane MM, Danaei G, Pelizzari PM, Cowan MJ, Paciorek CJ, et al. (2011) National, regional, and global trends in serum total cholesterol since 1980: systematic analysis of health examination surveys and epidemiological studies with 321 country-years and 3·0 million participants. The Lancet 377: 578–586. doi: 10.1016/S0140-6736(10)62038-7 PMID: 21295847

9. Cohen JD, Cziraky MJ, Cai Q, Wallace A, Wasser T, Crouse JR, et al. (2010) 30-year trends in serum lipids among United States adults: results from the National Health and Nutrition Examination Surveys II, III, and 1999–2006. Am J Cardiol 106: 969–975. doi: 10.1016/j.amjcard.2010.05.030 PMID: 20854959

10. Pietinen P, Vartiainen E, Seppänen R, Aro A, Puska P (1996) Changes in diet in Finland from 1972 to 1992: impact on coronary heart disease risk. Prev Med 25: 243–250. doi: 10.1006/pmed.1996.0053 PMID: 8781001

11. Johnson CL, Rifkind BM, Sempos CT, Carroll MD, Bachorik PS, Briefel RR, et al. (1993) Declining serum total cholesterol levels among US adults: The National Health and Nutrition Examination Surveys. JAMA 269: 3002–3008. doi: 10.1001/jama.1993.03500230084034 PMID: 8501842

12. Thorsson B, Steingrimsdottir L, Halldorsdottir S, Andersen K, Sigurdsson G, Aspelund T, et al. (2013) Changes in total cholesterol levels in Western societies are not related to statin, but rather dietary factors: the example of the Icelandic population. Eur Heart J 34: 1778–1782. doi: 10.1093/eurheartj/ehs395 PMID: 23209261

13. Eliasson M, Janlert U, Jansson JH, Stegmayr B (2006) Time trends in population cholesterol levels 1986–2004: influence of lipid-lowering drugs, obesity, smoking and educational level. The northern Sweden MONICA study. J Intern Med 260: 551–559. doi: 10.1111/j.1365-2796.2006.01730.x PMID: 17116006

14. Cífková R, Škodová Z, Bruthans J, Adámková V, Jozífová M, Galovcová M, et al. (2010) Longitudinal trends in major cardiovascular risk factors in the Czech population between 1985 and 2007/8. Czech MONICA and Czech post-MONICA. Atherosclerosis 211: 676–681. doi: 10.1016/j.atherosclerosis.2010.04.007 PMID: 20471016

15. Valsta LM, Tapanainen H, Sundvall J, Laatikainen T, Männistö S, Pietinen P, et al. (2010) Explaining the 25-year decline of serum cholesterol by dietary changes and use of lipid-lowering medication in Finland. Public Health Nutr 13: 932–938. doi: 10.1017/S1368980010001126 PMID: 20513263

16. Blackman T (2007) Statins, saving lives, and shibboleths. BMJ 334: 902. doi: 10.1136/bmj.39163.563519.55 PMID: 17463464

17. Capewell S, Graham H (2010) Will cardiovascular disease prevention widen health inequalities? PLoS Med 7: e1000320. doi: 10.1371/journal.pmed.1000320 PMID: 20811492

18. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. (2014) 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 129: S1–S45. doi: 10.1161/01.cir.0000437738.63853.7a PMID: 24222016

19. National Institute for Health and Care Excellence (2014) Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular

disease. NICE guidelines [CG181]. Available: http://www.nice.org.uk/Guidance/CG181. Accessed 26 August 2014.

20. Wise J (2014) Open letter raises concerns about NICE guidance on statins. BMJ 348: 3937. doi: 10.1136/bmj.g3937

21. Thompson R, Gerada C, Haslam D, Bamrah JS, Kendrick M, Malhotra A, et al. (2014) NICE statin letter: concerns about the latest NICE draft guidance on statins. Available: www.nice.org.uk/media/877/AC/NICE_statin_letter.pdf.

22. Department for Communities and Local Government (2011) English indices of deprivation 2010—Publications—GOV.UK. Available: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010. Accessed 26 August 2014.

23. NatCen Social Research, University College London. Department of Epidemiology and Public Health (2014) Health Survey for England, 2012 [computer file]. Colchester, Essex: UK Data Archive [distributor]. Available: http://dx.doi.org/10.5255/UKDA-SN-7480-1. Accessed 1 May 2014.

24. NatCen Social Research, University College London. Department of Epidemiology and Public Health (2013) Health Survey for England, 2011 [computer file]. Colchester, Essex: UK Data Archive [distributor]. Available: http://dx.doi.org/10.5255/UKDA-SN-7260-1. Accessed 1 May 2014.

25. Office of Population Censuses and Surveys. Social Survey Division (1997) Health Survey for England, 1991–1992: combined data file [computer file]. 2nd edition. Colchester, Essex: UK Data Archive [distributor]. Available: http://dx.doi.org/10.5255/UKDA-SN-3238-1. Accessed 1 May 2014.

26. Bridges S, Doyle M, Fuller E, Knott C, Mindell J, Moody A, et al. (2013) Health Survey for England 2012. Methods and documentation. Health and Social Care Information Centre. Available: http://www.hscic.gov.uk/catalogue/PUB13218/HSE2012-Methods-and-docs.pdf.

27. Boniface S, Bridges S, Craig R, Darton R, Fuller E, Hancock R, et al. (2012) Health Survey for England 2011. Methods and documentation. Health and Social Care Information Centre. Available: http://www.hscic.gov.uk/catalogue/PUB09300/HSE2011-Methods-and-docs.pdf.

28. UK Data Service (n.d.) Health Survey for England, 1991–1992: combined data file, documentation. Available: http://discover.ukdataservice.ac.uk/catalogue/?sn=3238&type=Data%20catalogue. Accessed 16 June 2014.

29. Office for National Statistics (2012) Standard Occupational Classification and Socio-Economic Classification: Archive. Available: http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/soc-and-sec-archive/index.html. Accessed 27 August 2014.

30. Capewell S, Ford ES (2011) Why have total cholesterol levels declined in most developed countries? BMC Public Health 11: 641. doi: 10.1186/1471-2458-11-641 PMID: 21834954

31. Capewell S, Morrison CE, McMurray JJ (1999) Contribution of modern cardiovascular treatment and risk factor changes to the decline in coronary heart disease mortality in Scotland between 1975 and 1994. Heart 81: 380–386. doi: 10.1136/hrt.81.4.380 PMID: 10092564

32. R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available: http://www.R-project.org/.

33. Lumley T (2013) Survey: analysis of complex survey samples. Available: http://cran.r-project.org/web/packages/survey/index.html.

34. Korn LE, Graubard IB (1998) Confidence Intervals for Proportions with Small Expected Number of Positive Counts Estimated from Survey Data. Surv Methodol 24: 193–201.

35. Law MR, Wald NJ, Rudnicka AR (2003) Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. BMJ 326: 1423. doi: 10.1136/bmj.326.7404.1423 PMID: 12829554

36. Kong SX, Crawford SY, Gandhi SK, Seeger JD, Schumock GT, Lam NP, et al. (1997) Efficacy of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors in the treatment of patients with hypercholesterolemia: a meta-analysis of clinical trials. Clin Ther 19: 778–797. doi: 10.1016/S0149-2918(97)80102-6 PMID: 9377621

37. Edwards JE, Moore RA (2003) Statins in hypercholesterolaemia: A dose-specific meta-analysis of lipid changes in randomised, double blind trials. BMC Fam Pract 4: 18. doi: 10.1186/1471-2296-4-18 PMID: 14969594

38. Jones MD P, Kafonek MD S, Laurora PharmD I, Hunninghake MD D (1998) Comparative dose efficacy study of atorvastatin versus simvastatin, pravastatin, lovastatin, and fluvastatin in patients with hypercholesterolemia (the CURVES study). Am J Cardiol 81: 582–587. doi: 10.1016/S0002-9149(97)00965-X PMID: 9514454

39. Health and Social Care Information Centre (2013) Prescription Cost Analysis 2012. Available: http://www.hscic.gov.uk/searchcatalogue?productid=11412&q = prescription+cost+analysis&topics=0%2fPrescribing&sort=Relevance&size=10&page=1#top. Accessed 23 February 2014.

40. Health and Social Care Information Centre 1 Trevelyan Square (2012) Prescription Cost Analysis 2011. Available: http://www.hscic.gov.uk/pubs/prescostanalysis2011. Accessed 23 February 2014.

41. Cochran WG (1977) Sampling Techniques. 3rd edition. New York: John Wiley & Sons. 428 p.

42. Gatz DF, Smith L (1995) The standard error of a weighted mean concentration—I. Bootstrapping vs other methods. Atmos Environ 29: 1185–1193. doi: 10.1016/1352-2310(94)00210-C

43. Ekstrom C (2014) MESS: Miscellaneous esoteric statistical script. Available: http://CRAN.R-project.org/package=MESS.

44. Bajekal M, Scholes S, Love H, Hawkins N, O'Flaherty M, Raine R, et al. (2012) Analysing recent socio-economic trends in coronary heart disease mortality in England, 2000–2007: a population modelling study. PLoS Med 9: e1001237. doi: 10.1371/journal.pmed.1001237 PMID: 22719232

45. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P (2007) Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 335: 136. doi: 10.1136/bmj.39261.471806.55 PMID: 17615182

46. Collins GS, Altman DG (2012) Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 344: e4181–e4181. doi: 10.1136/bmj.e4181 PMID: 22723603

47. Ashworth M, Lloyd D, Smith RS, Wagner A, Rowlands G (2007) Social deprivation and statin prescribing: a cross-sectional analysis using data from the new UK general practitioner "Quality and Outcomes Framework." J Public Health 29: 40–47. doi: 10.1093/pubmed/fdl068 PMID: 17071815

48. Wu J, Zhu S, Yao GL, Mohammed MA, Marshall T (2013) Patient factors influencing the prescribing of lipid lowering drugs for primary prevention of cardiovascular disease in UK general practice: a national retrospective cohort study. PLoS ONE 8: e67611. doi: 10.1371/journal.pone.0067611 PMID: 23922649

49. Stocks NP, Ryan P, McElroy H, Allan J (2004) Statin prescribing in Australia: socioeconomic and sex differences. A cross-sectional study. Med J Aust 180: 229–231. PMID: 14984343

50. Ohlsson H, Lynch K, Merlo J (2010) Is the physician's adherence to prescription guidelines associated with the patient's socio-economic position? An analysis of statin prescription in South Sweden. J Epidemiol Community Health 64: 678–683. doi: 10.1136/jech.2008.081166 PMID: 19692716

51. Thomsen RW, Johnsen SP, Olesen AV, Mortensen JT, Bøggild H, Olsen J, et al. (2005) Socioeconomic gradient in use of statins among Danish patients: population-based cross-sectional study. Br J Clin Pharmacol 60: 534–542. doi: 10.1111/j.1365-2125.2005.02494.x PMID: 16236044

52. Franks P, Tancredi D, Winters P, Fiscella K (2010) Cholesterol treatment with statins: Who is left out and who makes it to goal? BMC Health Serv Res 10: 68. doi: 10.1186/1472-6963-10-68 PMID: 20236527

53. Shekelle RB, Shryock AM, Paul O, Lepper M, Stamler J, Liu S, et al. (1981) Diet, serum cholesterol, and death from coronary heart disease: the Western Electric study. N Engl J Med 304: 65–70. doi: 10.1056/NEJM198101083040201 PMID: 7442730

54. Food and Agriculture Organization of the United Nations (2014) FAOSTAT. Available: http://faostat.fao.org/default.aspx. Accessed 16 June 2014.

55. Rose G (2001) Sick individuals and sick populations. Int J Epidemiol 30: 427–432. doi: 10.1093/ije/30.3.427 PMID: 11416056

56. Ellis JJ, Erickson SR, Stevenson JG, Bernstein SJ, Stiles RA, Fendrick MA (2004) Suboptimal statin adherence and discontinuation in primary and secondary prevention populations. J Gen Intern Med 19: 638–645. doi: 10.1111/j.1525-1497.2004.30516.x PMID: 15209602

57. Chaudhry HJ, McDermott B (2008) Recognizing and improving patient nonadherence to statin therapy. Curr Atheroscler Rep 10: 19–24. PMID: 18366981

58. Bouchard M-H, Dragomir A, Blais L, Berard A, Pilon D, Perreault S (2007) Impact of adherence to statins on coronary artery disease in primary prevention. Br J Clin Pharmacol 63: 698–708. doi: 10.1111/j.1365-2125.2006.02828.x PMID: 17214831

59. Wallach-Kildemoes H, Andersen M, Diderichsen F, Lange T (2013) Adherence to preventive statin therapy according to socioeconomic position. Eur J Clin Pharmacol 69: 1553–1563. doi: 10.1007/s00228-013-1488-6 PMID: 23588558

60. Sugiyama T, Tsugawa Y, Tseng C-H, Kobayashi Y, Shapiro MF (2014) Different time trends of caloric and fat intake between statin users and nonusers among US adults: gluttony in the time of statins? JAMA Intern Med 174: 1038–1045. doi: 10.1001/jamainternmed.2014.1927 PMID: 24763487

61. Jørgensen T, Capewell S, Prescott E, Allender S, Sans S, Zdrojewski T, et al. (2013) Population-level changes to promote cardiovascular health. Eur J Prev Cardiol 20: 409–421. doi: 10.1177/2047487312441726 PMID: 22514213

62. Brunner EJ, Marmot MG, White IR, O'Brien JR, Etherington MD, Slavin BM, et al. (1993) Gender and employment grade differences in blood cholesterol, apolipoproteins and haemostatic factors in the Whitehall II study. Atherosclerosis 102: 195–207. doi: 10.1016/0021-9150(93)90162-N PMID: 8251006

63. Gazmararian JA, Kripalani S, Miller MJ, Echt KV, Ren J, Rask K (2006) Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. J Gen Intern Med 21: 1215–1221. doi: 10.1111/j.1525-1497.2006.00591.x PMID: 17105519

64. Hippisley-Cox J, Parker C, Coupland C, Vinogradova Y (2007) Inequalities in the primary care of patients with coronary heart disease and serious mental health problems: a cross-sectional study. Heart 93: 1256–1262. doi: 10.1136/hrt.2006.110171 PMID: 17344333

Cardiovascular screening to reduce the burden from cardiovascular disease: microsimulation study to quantify policy options

Chris Kypridemos,[1] Kirk Allen,[1, 2] Graeme L Hickey,[3] Maria Guzman-Castillo,[1] Piotr Bandosz,[1, 4] Iain Buchan,[5] Simon Capewell,[1] Martin O'Flaherty[1]

[1]Department of Public Health and Policy, University of Liverpool, Liverpool L69 3GB, UK

[2]Lancaster Medical School, Lancaster University, Lancaster, UK

[3]Department of Biostatistics, University of Liverpool, Liverpool, UK

[4]Department of Prevention and Medical Education, Medical University of Gdansk, Gdansk, Poland

[5]Farr Institute @ HeRC, University of Manchester, Manchester, UK

Correspondence to:
C Kypridemos
c.kypridemos@liverpool.ac.uk

ABSTRACT

OBJECTIVES
To estimate the potential impact of universal screening for primary prevention of cardiovascular disease (National Health Service Health Checks) on disease burden and socioeconomic inequalities in health in England, and to compare universal screening with alternative feasible strategies.

DESIGN
Microsimulation study of a close-to-reality synthetic population. Five scenarios were considered: baseline scenario, assuming that current trends in risk factors will continue in the future; universal screening; screening concentrated only in the most deprived areas; structural population-wide intervention; and combination of population-wide intervention and concentrated screening.

SETTING
Synthetic population with similar characteristics to the community dwelling population of England.

PARTICIPANTS
Synthetic people with traits informed by the health survey for England.

MAIN OUTCOME MEASURE
Cardiovascular disease cases and deaths prevented or postponed by 2030, stratified by fifths of socioeconomic status using the index of multiple deprivation.

RESULTS
Compared with the baseline scenario, universal screening may prevent or postpone approximately 19 000 cases (interquartile range 11 000-28 000) and 3000 deaths (−1000-6000); concentrated screening 17 000 cases (9000-26 000) and 2000 deaths (−1000-5000); population-wide intervention 67 000 cases (57 000-77 000) and 8000 deaths (4000-11 000); and the combination of the population-wide intervention and concentrated screening 82 000 cases (73 000-93 000) and 9000 deaths (6000-13 000). The most equitable strategy would be the combination of the population-wide intervention and concentrated screening, followed by concentrated screening alone and the population-wide intervention. Universal screening had the least apparent impact on socioeconomic inequalities in health.

CONCLUSIONS
When primary prevention strategies for reducing cardiovascular disease burden and inequalities are compared, universal screening seems less effective than alternative strategies, which incorporate population-wide approaches. Further research is needed to identify the best mix of population-wide and risk targeted CVD strategies to maximise cost effectiveness and minimise inequalities.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Two main strategies for the primary prevention of cardiovascular disease (CVD) is to screen the population, find those individuals at high risk, and treat them or to reduce the CVD risk of the whole population irrespective of individuals' baseline risk

Evidence suggests that the second approach is more effective and likely more equitable, yet this depends on the distribution of CVD risk throughout the population

In England, the Department of Health adopted the first approach, although this decision has recently attracted some criticism

WHAT THIS STUDY ADDS

In England, despite the observed higher concentration of CVD risk in more deprived areas, structural population-wide interventions targeting unhealthy diet and tobacco might be three times more effective than the existing screening policy

Structural population-wide interventions are also likely to be more equitable than screening

A comprehensive strategy, combining structural population-wide interventions with screening in the most deprived areas (where CVD risk is concentrated) is most likely to maximise both effectiveness and equity of primary CVD prevention

Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide.[1] Furthermore, substantial socioeconomic inequalities have been observed in CVD mortality in England and elsewhere.[2 3] These inequalities powerfully reflect much greater premature mortality, and hence shorter life expectancy, among the most deprived groups. In England, the current governmental action plan to tackle the burden of CVD includes a programme known as NHS (National Health Service) Health Checks. Introduced in 2009, this programme promotes the screening of all healthy adults aged 40 to 74 for CVD risk stratification, and treatment of those at high risk.[4 5] Recently, the debate about the programme's scientific foundation, effectiveness, and cost effectiveness, however, has been heated.[6-10] Despite the controversy, the programme remains policy.

Beyond the obvious importance of the debate to national public health, the programme's relevance extends internationally. Choices about public health policy in the United Kingdom influence policy worldwide; the UK policies on tobacco control and salt reduction are two recent examples.[11 12] In essence, the debate about NHS Health Checks originates from the archetypal debate of targeted "high risk" versus "population-wide" preventive interventions that was first articulated by Geoffrey Rose.[13] Rose argued that

population-wide interventions are more effective than ones aimed at high risk groups because the majority of incident cases occur in the multitudinous group of people at low and intermediate risk. In Rose terminology, NHS Health Checks is a typical "high risk intervention," as it targets people at high risk rather than lowering risk in the whole population.

The effectiveness of high risk interventions for CVD prevention has been previously challenged.[14] More recently, a Cochrane systematic review and the Inter99 trial found no benefits of health checks on CVD morbidity or mortality.[15 16] There were, however, major limitations to these studies: Inter99 trialled a counselling intervention not supported by additional drug treatment, and in the Cochrane review nine out of 14 trials were conducted before 1980, when the treatment options for high risk people were limited. In addition, high risk interventions may be more effective in populations with high clustering of risk factors, resulting in a high concentration of the risk to certain groups in the population.[17] In fact, the English population has such characteristics, with the risk of CVD being higher among those in the most socioeconomically deprived groups.[18]

High risk interventions may generate health inequalities because they require active participation of people in both screening and treatment of those at high risk, favouring those with more resources.[14 19-21] The particular effect of NHS Health Checks on socioeconomic health inequalities remains unclear however. A national study reported no difference in the coverage of the intervention by deprivation,[22] whereas several smaller, but more detailed, studies showed substantially lower uptake in deprived areas.[23-25]

We estimated the potential impact of universal screening for primary prevention of CVD on disease burden and socioeconomic health inequalities in England. Available data on the effectiveness of the NHS Health Check programme have been used to model this scenario. We further compared universal CVD screening with an alternative approach targeting only deprived areas, a feasible population-wide intervention, and a combination of both.

Methods

Building on experience from the original, validated IMPACT model[26] and the more recent IMPACT$_{SEC}$[27] and IMPACT2 models,[28] we created IMPACT$_{NCD}$, a discrete time dynamic stochastic microsimulation model. IMPACT$_{NCD}$ simulates the life course of synthetic individuals under different counterfactual scenarios, up to 2030 (the projection horizon). During the simulation, CVD incidence and CVD and non-CVD mortality are recorded. The results are stratified by year, five year age group, sex, and fifths of index of multiple deprivation. The last is a relative measure of area deprivation that is widely used by public health authorities in England, and it has been used as the measure of socioeconomic classification for this study.[29]

A more detailed description of the model is provided in the supplementary material and the source code is available at https://github.com/ChristK/IMPACTncd/tree/CVD-policy-options.

Scenarios
We considered five scenarios.

Baseline (current trends)
In the baseline scenario, we assumed that the recent observed trends in CVD risk factor trajectories by age, sex, and socioeconomic status will continue in the near future. We extracted the trends from the health survey for England 2001-12, a nationally representative series of health surveys conducted in England annually.[30-42]

Universal screening
This scenario modelled the potential health effects of universal screening to identify and treat people at high risk for CVD. Input variables were informed from current implementation of the NHS Health Check programme. Eligible people were defined as adults aged between 40 and 74, excluding those with a known history of CVD, atrial fibrillation, diabetes mellitus, rheumatoid arthritis, or renal disease; closely resembling real life eligibility criteria. Based on existing evidence we assumed an uptake of 50% for screening,[43] and we calibrated the distribution of the estimated 10 year risk of developing CVD among those participating: 70% with a less than 10% risk, 25% with between 10% and 20%, and 5% with more than 20%.[22] In addition, we calibrated the age distribution so that around 30% of those screened were older than 60.[22] Participants with a higher than 10% estimated 10 year risk of developing CVD were considered at high risk and eligible for treatment. We used the QRISK2 score to estimate the 10 year risk of developing CVD, as perceived from healthcare.[44]

Based on published evidence, we assumed that about 24% with an estimated risk of 20% or more and total cholesterol of 5 mmol/L or more will be prescribed atorvastatin 20 mg and about 27% with an estimated risk of 20% or more and a systolic blood pressure of 135 mm Hg or more will be prescribed antihypertensive drugs. For those with a risk between 10% and 20% we assumed that about 17% and 20% will be prescribed treatment, respectively.[45] We assumed an 80% persistence with treatment and a mean adherence of approximately 70%, roughly based on evidence from Denmark.[46] Moreover, we modelled high risk participants with a body mass index of more than 50 kg/m² to undergo bariatric surgery and reduce their body mass index to 30 kg/m². We assumed that with lifestyle counselling half of the high risk participants consuming fewer than five fruit and vegetable portions daily will increase their consumption by a portion daily. Half of those being active for less than five days a week will increase their physical activity by an active day each week, and all high risk participants will decrease their body mass index by around 1%.[45 47] Finally, we modelled 10% of high risk smokers to achieve cessation for a year and have a probability of relapse equal to that of the general population by sex, fifth of multiple deprivation, and years since cessation.[48 49]

Concentrated screening

In the concentrated screening scenario, we simulated a hypothetical strategy where screening had only been implemented in the most deprived fifths (groups 4 and 5), the groups with the greatest concentration of CVD risk. We assumed that the uptake of the intervention was 50% and the risk and age distribution in the participants was similar to that in the eligible population. Otherwise, the strategy is similar to the previous universal screening scenario. Given the recent criticism about the cost and cost effectiveness of the intervention,[9] offering the intervention where the risk is more concentrated may reduce costs.

Population-wide intervention

This scenario modelled the effects of a feasible population-wide structural intervention targeting unhealthy diet and smoking. Several studies have found that a tax on sugar sweetened beverages may reduce the prevalence of obesity.[50-52] For this scenario we assumed that such a tax may reduce the mean increase in body mass index by about 5% annually. Moreover, the United Kingdom has had one of the world's most successful salt reduction strategies, including public awareness campaigns, food labelling, and voluntary reformulation of processed foods.[53] Modelling studies suggested that the addition of mandatory reformulation of processed foods may further reduce mean systolic blood pressure by 0.8 mm Hg;[54] we modelled this decrease. A large randomised trial in the United States showed that subsidies on fruits and vegetables may increase consumption by about half a portion daily, and a modelling study in the UK found that subsidising fruits and vegetables combined with taxation of unhealthy foods may increase fruit and vegetable annual consumption by about 10%.[55 56] We modelled an increase of a portion of fruit and vegetable each day in 50% of the population. Finally, a SimSmoke modelling study estimated that full compliance with the framework convention on tobacco control may reduce smoking prevalence by 13% (relative) in five years;[57] we modelled this decrease.

Population-wide intervention and concentrated screening

This scenario is the combination of the population-wide intervention and concentrated screening strategies. We modelled the implementation of a population-wide strategy identical to the previous scenario, complemented by concentrated screening for people at high risk of CVD in the most deprived fifths (groups 4 and 5).

Common scenario assumptions

All interventions begun in 2011 and were linearly diffused into the population over a five year period. Trends in population risk factors were assumed to be the same as those of the baseline scenario for all but the population-wide intervention. All of the scenarios assumed that CVD case fatality will keep improving by 3% (relative) annually. In addition, we assumed a socioeconomic gradient in CVD case fatality, forcing the more deprived people to experience worse outcomes. Both case fatality assumptions were based on recent trends and are supported by the British Heart Foundation's statistics on coronary heart disease.[2] Finally, a five year lag time was assumed between exposure to cardiovascular risk factors and disease.

Model description

Inputs and logic

IMPACT$_{NCD}$ synthesises information from the Office for National Statistics and the health surveys for England on the English population's demographics and its exposure to CVD associated risk factors, to generate a close-to-reality synthetic population.[58] Well established causal pathways between CVD and the associated risk factors are used to translate exposure into CVD incidence and mortality, in a competing risk framework. We obtained effect sizes for exposures from published meta-analyses and longitudinal studies (see supplementary table S1).

The risk factors we considered for this study were age, sex, fifth of deprivation, body mass index, systolic blood pressure, total cholesterol level, diabetes mellitus (diagnosis or increased glycated haemoglobin level/no diabetes), smoking status (current, former, or never smoker), environmental tobacco exposure (binary variable), fruit and vegetable consumption (portions daily), and physical activity (days with at least 30 minutes of moderate or vigorous physical activity each week). CVD was defined as the sum of coronary heart disease and stroke (any type) cases. As this study focuses on primary prevention, we considered only the first ever episode of coronary heart disease or stroke. The competing risk framework allowed people to develop coronary heart disease and/or stroke separately, and to die from these two diseases or any other cause.

Model outputs

We report the cumulative estimates of cases and deaths prevented or postponed as measures of overall effectiveness of the modelled interventions. To measure the impact of the modelled interventions on absolute and relative socioeconomic health inequalities, we developed and used two regression based metrics inspired by the slope index of inequality;[59] the absolute equity slope index and the relative equity slope index. The absolute equity slope index measures the impact of an intervention on absolute inequality; for example, a value of 100 means 100 more cases were prevented or postponed in most deprived areas compared with least deprived areas, resulting in a decrease in absolute inequality. The relative equity slope index takes into account the pre-existing socioeconomic gradient of disease burden and measures the impact of an intervention on relative inequality. Positive values mean the intervention tackles relative inequalities and negative values that the intervention generates relative inequality. Finally, we summarised the overall impact of each scenario on CVD burden and equity in the equity summary chart.

Uncertainty and sensitivity analysis

IMPACT$_{NCD}$ implements a second order Monte Carlo design that allows uncertainty to be quantified from the outputs. We used distributions to model the uncertainty around all scenario specific inputs and the sampling error of the risk associated with the CVD related risk factors. The probabilistic sensitivity analysis has been incorporated in our estimates. We summarise the distributions by reporting medians and interquartile ranges in the form of first and third fourths. The supplementary file provides a more detailed description of the sources of uncertainty and the relevant distributions.

We ran three further scenarios offering slight variations on the two primary ones of universal screening and population-wide intervention: a universal screening variation, where we assumed a treatment threshold recommendation of 20% risk instead of 10%; another variation on universal screening, where we assumed a socioeconomic differential in screening uptake, with the most deprived of the population to be 10% less likely to participate; and a variation on the population-wide intervention, where we only modelled dietary interventions, excluding smoking interventions. The supplementary file provides detailed information on the extra scenarios.

Validation

We assessed the predictive validity of the IMPACT$_{NCD}$ model by comparing the estimated number of deaths from CVD with the observed number of deaths from the same causes for 2006 to 2013 in England.[60] We further compared the IMPACT$_{NCD}$ output with CVD mortality forecasts from a bayesian age-period-cohort model.[61]

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

IMPACT$_{NCD}$ outputs for CVD burden and inequality are summarised for ages 30 to 84. Because of the assumed five year time lag, the interventions affect the population from 2016 up to the projection horizon of 2030. The impact of the five scenarios on risk factor trajectories are further illustrated in additional graphs in the supplementary file.

Overall effectiveness

Under the baseline scenario, IMPACT$_{NCD}$ estimated about 1.4 million (interquartile range 1.3-1.5) cases of CVD and 540 000 deaths (interquartile range 520 000 to 550 000) between 2016 and 2030. The most effective intervention was the combination of the population-wide intervention and concentrated screening. The population-wide intervention alone had the second highest effectiveness, whereas the universal and the concentrated screening scenarios were considerably less effective (table 1). Despite the improvement of most CVD related risk factors, the proportion of high risk people in the eligible population is slowly increasing over time, because of population aging (fig 1).

Socioeconomic inequalities

When socioeconomic inequalities were considered, the patterns for reductions in absolute and relative inequalities were similar. The combination of the population-wide intervention and concentrated screening seemed the most powerful among the simulated interventions (tables 2 and 3). Concentrated screening alone was the second most powerful intervention in tackling inequalities, followed by the population-wide intervention. Finally, universal screening of CVD is likely to have a small, if any, effect on socioeconomic inequalities.

Equity summary chart

We summarised our estimates for the effectiveness and equity of the modelled interventions in the equity summary chart (fig 2). The horizontal axis of the chart represents the cases of CVD prevented or postponed and the vertical axis the reduction in absolute inequality. Scenarios above the equity curve (dashed curve in the figure) decrease relative socioeconomic inequality, and scenarios below the curve increase it. The vertical distance from the curve approximates the impact of the scenario on relative inequality. (See the supplementary file for more details about this chart.) The combination of the population-wide intervention and concentrated screening is by far the most effective and equitable intervention. Concentrated screening is also equitable but with few mortality gains.

Sensitivity analysis

Adding assumptions to extend the scenarios did not displace our main findings. The three most notable results of the sensitivity analysis were:

Raising the treatment threshold from 10% to 20% further reduced the effectiveness of universal screening by about 60% in preventing CVD cases. However, in preventing deaths from CVD the effectiveness decreased by only 15% as raising the treatment threshold excludes younger participants at intermediate risk from treatment.

Assuming a differential uptake of universal screening by deprivation fifth essentially eliminated the estimated small potential benefit of universal screening in tackling health inequalities.

A population-wide intervention targeting only diet would still be about twice as effective as universal

Table 1 | Estimated cases and deaths prevented or postponed under each scenario, by 2030

Scenarios	No (interquartile range) prevented or postponed	
	Cases	Deaths
Universal screening	19 000 (11 000-28 000)	3000 (−1000-6000)
Concentrated screening	17 000 (9000-26 000)	2000 (−1000-5000)
Population-wide intervention	67 000 (57 000-77 000)	8000 (4000-11 000)
Population-wide intervention and concentrated screening	82 000 (73 000-93 000)	9000 (6000-13 000)

Results rounded to nearest 1000.

Fig 1 | Proportion of high risk people eligible for universal screening population projections, by age group and sex. 10 year risk of cardiovascular disease (CVD) was estimated from QRISK2 score. Error bars represent interquartile ranges

screening and more than twice as effective as population-wide intervention targeting smoking alone—so the relative ranking of scenario effectiveness would remain unaltered. For detailed results see supplementary tables S11-S13.

Validation

We assessed the predictive validity of the IMPACT$_{NCD}$ model by comparing the estimated number of deaths from CVD with the observed number of deaths from the same cause for 2006 to 2013 in England (fig 3). See the supplementary file for detailed graphs by age group, sex, deprivation fifth, and disease.

Discussion

Our results strongly suggest that universal screening and treatment of people at high risk is not the most effective option for primary prevention of cardiovascular disease (CVD) overall, nor for reducing socioeconomic inequalities. In contrast, prevention

Table 2 | Cases prevented or postponed according to fifth of deprivation by 2030, along with absolute equity slope index for each scenario

Deprivation fifth*	No (interquartile range) of cases prevented or postponed			
	Universal screening	Concentrated screening	Population-wide intervention	Population-wide intervention+ concentrated screening
First (least deprived)	3400 (−1400-8300)	0	10 800 (5900-15 500)	10 800 (6200-15 700)
Second	2900 (−1500-8400)	0	12 200 (6200-17 200)	11 500 (6600-17 000)
Third	4000 (−900-9300)	0	13 100 (8100-18 300)	12 600 (7400-17 700)
Fourth	3700 (−1600-8600)	6400 (1500-11 800)	12 500 (7100-18 400)	18 700 (13 900-24 200)
Fifth (most deprived)	4900 (−600-10 400)	10 700 (5300-16 300)	18 700 (13 000-24 000)	28 600 (22 800-33 200)
Absolute equity slope index	1700 (−6200-9300)	14 100 (5700-23 000)	8400 (−400-16 900)	21 100 (12 800-29 300)

Results rounded to nearest 1000.
*According to index of multiple deprivation.

Table 3 | Relative percentage reduction in cases of cardiovascular disease according to fifth of deprivation by 2030, along with relative equity slope index for each scenario

Deprivation fifth*	Relative % reduction (interquartile range)			
	Universal screening	Concentrated screening	Population-wide intervention	Population-wide intervention+ concentrated screening
First (least deprived)	1.3 (−0.5-3.1)	0	4.1 (2.2-5.9)	4.0 (2.4-6.0)
Second	1.1 (−0.5-2.9)	0	4.2 (2.2-5.9)	4.0 (2.3-5.9)
Third	1.4 (−0.3-3.2)	0	4.6 (2.8-6.3)	4.4 (2.6-6.2)
Fourth	1.3 (−0.6-3.1)	2.4 (0.6-4.3)	4.6 (2.7-6.6)	6.9 (5.1-8.9)
Fifth (most deprived)	1.6 (−0.2-3.3)	3.6 (1.8-5.3)	6.2 (4.4-8.0)	9.4 (7.6-11.2)
Relative equity slope index	0.4 (−2.4-3.2)	4.9 (1.8-7.9)	2.3 (−0.7-5.3)	6.7 (3.8-9.5)

Results rounded to one decimal place.
*According to index of multiple deprivation.

strategies that include population-wide structural interventions seem to be the consistently better options for reducing overall CVD burden and inequalities. This echoes and quantifies findings from other, mostly theoretical, studies supporting that structural population-wide interventions are powerful, while reducing socioeconomic health inequalities.[13 14 62 63] Indeed, the impact of the population-wide intervention scenario on reduction in estimated mortality and inequalities seems compatible with previous estimates, considering the different methodologies.[64] Furthermore, the effectiveness and equity of population-wide structural interventions can be further improved by the addition of targeted interventions in the most deprived groups, as highlighted in the combined scenario of the population-wide intervention and concentrated screening.

Compared with other modelling approaches, our IMPACT$_{NCD}$ model estimated that NHS Health Checks might prevent approximately 1000 non-fatal and 200 fatal cases of CVD annually. This is comparable with the Department of Health estimates of 1600 non-fatal CVD cases and 650 deaths prevented annually.[4] Furthermore, the Department of Health modelling approach assumed an intervention uptake of 75%; higher than the current observed levels. Using the Archimedes model, Schuetz et al estimated that health checks in the UK could prevent some 12 CVD cases per 1000 population screened after 30 years' follow-up[65] (7500 CVD cases prevented each year extrapolating to the eligible English population). That higher estimate reflects the researchers' apparently unrealistic assumption of 100% screening uptake and 50% overall uptake of treatment.

The scenarios

We modelled the universal screening scenario to closely resemble the current implementation of the NHS Health Check programme, based on published evidence. Therefore, we maintain that our estimates on the effectiveness of this scenario are not far from the real world effectiveness of NHS Health Checks. However, our output suggesting that universal screening might reduce socioeconomic inequalities seems to contradict existing empirical and modelling evidence.[14 19-21] This is because we generously assumed identical screening uptake and treatment adherence for all socioeconomic groups. In fact, any potential reduction in socioeconomic health inequalities was essentially eliminated when we considered a small socioeconomic differential in uptake in the sensitivity analysis. Furthermore, additional health inequalities may arise from differential persistence and adherence to treatment by deprivation status.[46]

The population-wide intervention scenario on the other hand, is based mostly on structural policies targeting price and availability. This scenario potential effectiveness was mostly based on natural experiments,[66 67] and on previous modelling studies from the UK and elsewhere. The size of the changes in the population risk factors that we modelled were modest, and actually smaller than the reductions observed in countries such as France, Finland, and the USA during recent decades.[68-70] This scenario estimated the reduction in mortality conservatively, because it ignored the beneficial effect of the policies on survival from CVD. Similarly, it underestimated the reduction of the gap in inequalities, because it did not fully consider the current disproportionate burden of poor diet among the
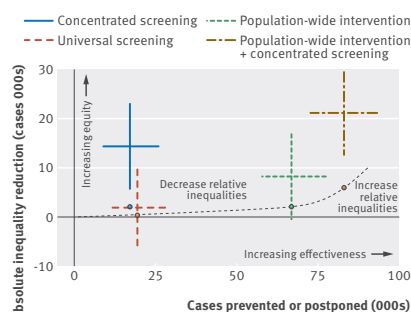
Fig 2 | Equity summary chart of effectiveness and equity of all modelled interventions, compared with baseline scenario (beginning of axes). Dashed line represents "equity" curve. Interventions below the curve increase relative inequality, whereas interventions above it decrease relative inequalities. Smaller coloured dots represent reference points used to fit equity curve. Horizontal and vertical error bars represent interquartile ranges

most deprived of the population,[71] and hence the potential for improvement through population-wide policies.

Finally, the concentrated screening strategy was the weakest in terms of overall effectiveness, yet more powerful in tackling inequalities. Its increased impact on socioeconomic health inequalities is a direct consequence of the concentrated prevention only to the more

Fig 3 | Number of deaths from cardiovascular disease (CVD) in England, by year for ages 30 to 84. Office for National Statistics reported deaths (observed) versus IMPACT$_{NCD}$ estimated. Observed deaths after 2010 were adjusted to account for changes in ICD-10 version used by the Office for National Statistics from 2011 onwards. Error bars represent interquartile ranges

deprived quantiles of the population. However, the scenario assumptions may not fully hold in real world implementation. Hence, concentrated screening represents a challenge for public health practitioners and policymakers to exploit the opportunity of a smaller and more homogeneous eligible population and to implement better recruitment and tactics for treatment adherence. Yet, cost effectiveness might also fall because of loss of economies of scale.

Public health implications

This IMPACT$_{NCD}$ modelling may help stakeholders to understand better the interplay between preventive policies, risk factors, disease, and inequalities, and thus potentially inform health policy and strategy. Hence, when compared with the alternative feasible interventions, universal screening seemed inferior both in primary prevention and in reducing socioeconomic health inequalities. Additionally, we estimated that the proportion of young people at high risk aged less than 60 in the eligible population will decrease in future (fig 1). This will render universal screening less effective and less cost effective for this age group, because a larger number will need to be screened to identify each high risk individual.

Our study suggests that despite the high clustering of risk factors in the most deprived parts of the population, structural population-wide approaches remain more effective than high risk ones for the prevention of CVD. Population-wide approaches also seem to be more effective in reducing absolute and relative socioeconomic health inequalities, generally cost much less than a universal screening programme, and may even be cost saving.[72 73] In this study, we did not model the full potential of these policies, as we focused only on diet and smoking interventions; we did not, for example, incorporate alcohol consumption or physical activity. In addition, we did not simulate the likely wider benefits of improved diet and smoking cessation on the plethora of relevant non-communicable diseases. Despite this restricted scope, for CVD prevention we estimated that structural policies targeting diet could be twice as effective as those targeting smoking. Yet, structural interventions for a healthier diet are currently underutilised compared with tobacco control. Several countries have now introduced taxes on sugary drinks or sugar, including Finland, France, Latvia, and Mexico. The UK has recently followed their example. Hungary is the only European country currently taxing unhealthy "junk" food.[74] However, fiscal interventions may face opposition from commercial vested interests.[75] Interestingly, an increasing body of evidence from empirical studies and modelling analyses suggest that the maximum health impact with a neutral effect on poverty may occur when food or drinks taxes are combined with subsidies for healthy foods.[56 76 77]

Moreover, the combination of a population-wide intervention with an intervention targeting the most deprived members, may further improve effectiveness and equity. This approach is in the spirit of proportionate universalism that was identified in the Marmot
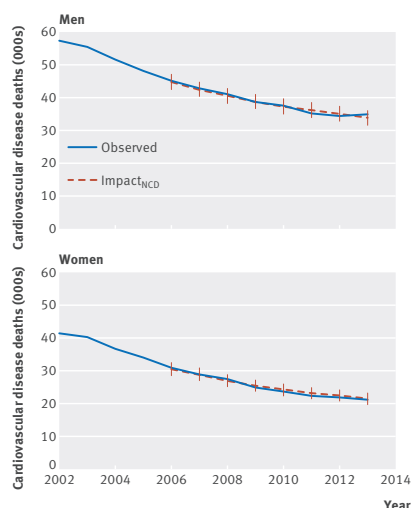
review as the best approach to tackle socioeconomic inequalities in health.[78] Our study provides evidence that in CVD prevention proportionate universalism may be the best option not only for tackling inequalities but also for overall effectiveness.

Strengths and limitations of this study

IMPACT$_{NCD}$ is the first microsimulation model to synthesise core principles of social and CVD epidemiology, vital demographics, published literature, and recent health surveys for England to create a synthetic population of England, including socioeconomic structure, at the individual level. The microsimulation approach allows for the simulation of detailed scenarios and explores the distributional nature of their impact on the population, in a competing risks framework. Microsimulation allows for greater flexibility and more detailed simulation, demanding more statistical and computational resources than older approaches; we utilised the Farr Institute's statistical high performance computing facilities.[79] Many assumptions must be made with such models. Yet, despite the potential frailty of such assumptions, this model validated well against observed CVD mortality, even when multiply stratified. Finally, to ensure transparency, we have made the IMPACT$_{NCD}$ source code open under GNU GPLv3 license.

Models are simplifications of reality and thus possess inherent limitations. At least four items were not included in the current model. Firstly, the multiplicative risk assumption is considered the status quo in comparative risk assessments;[80] however, this may oversimplify the complex nature of interactions between multiple risk factors and disease outcome over the life course. Secondly, IMPACT$_{NCD}$ currently ignores the effect of risk factors on CVD case fatality, although in this study we considered only primary prevention scenarios. Thirdly, complex population dynamics such as migration, social mobility, and the socioeconomic consequences of disease were not modelled. We consider this bias would be relatively small for projections with a short horizon. Fourthly, the model ignores the impact of universal screening in recognising previously undiagnosed cases of atrial fibrillation and other opportunistic diagnoses. Reassuringly, most of these biases apply across all scenarios; their effects would thus be reduced in comparisons between scenarios.

Conclusions

When comparing primary prevention strategies for reducing CVD burden and inequalities, universal screening seems less effective than alternative strategies that incorporate population-wide approaches. Further research is needed to identify the best mix of population-wide and risk targeted CVD strategies to maximise cost effectiveness and minimise inequalities.

1   GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;385:117-71. doi:10.1016/S0140-6736(14)61682-2.
2   Townsend N, Wickramasinghe K, Bhatnagar P, et al. Coronary heart disease statistics 2012 edition. British Heart Foundation 2012. www.bhf.org.uk/publications/view-publication.aspx?ps=1002097.
3   Mackenbach JP, Cavelaars AE, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. *Eur Heart J* 2000;21:1141-51. doi:10.1053/euhj.1999.1990.
4   Department of Health. Cardiovascular disease outcomes strategy: improving outcomes for people with or at risk of cardiovascular disease. 2013. www.gov.uk/government/uploads/system/uploads/attachment_data/file/217118/9387-2900853-CVD-Outcomes_web1.pdf.
5   Choices NHS. NHS Health Check. www.healthcheck.nhs.uk/.
6   Capewell S, McCartney M, Holland W. Invited debate: NHS Health Checks—a naked emperor? *J Public Health* 2015;37:187-92. doi:10.1093/pubmed/fdv063.
7   Waterall J, Greaves F, Kearney M, et al. Invited debate: NHS Health Check: an innovative component of local adult health improvement and well-being programmes in England. *J Public Health* 2015;37:177-84. doi:10.1093/pubmed/fdv062.
8   Braillon A, Bewley S, Pisinger C, Fisken RA, Richmond C. NHS health checks are a waste of resources. *BMJ* 2015;350:h1006. doi:10.1136/bmj.h1006.
9   O'Dowd A. MPs call for review of NHS health checks. *BMJ* 2014;349:g6498-6498. doi:10.1136/bmj.g6498.
10   Department of Health. Government response to the House of Commons Science and Technology Committee report on national health screening. 2015. www.gov.uk/government/publications/national-health-screening-government-response.
11   He FJ, Brinsden HC, MacGregor GA. Salt reduction in the United Kingdom: a successful experiment in public health. *J Hum Hypertens* 2014;28:345-52. doi:10.1038/jhh.2013.105.
12   Levy DT, Currie L, Clancy L. Tobacco control policy in the UK: blueprint for the rest of Europe? *Eur J Public Health* 2013;23:201-6. doi:10.1093/eurpub/cks090.
13   Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001;30:427-32, discussion 433-4. doi:10.1093/ije/30.3.427.

14 Capewell S, Graham H. Will cardiovascular disease prevention widen health inequalities? *PLoS Med* 2010;7:e1000320. doi:10.1371/journal.pmed.1000320.

15 Krogsbøll LT, Jørgensen KJ, Grønhøj Larsen C, Gøtzsche PC. General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. *BMJ* 2012;345:e7191. doi:10.1136/bmj.e7191.

16 Jørgensen T, Jacobsen RK, Toft U, Aadahl M, Glümer C, Pisinger C. Effect of screening and lifestyle counselling on incidence of ischaemic heart disease in general population: Inter99 randomised trial. *BMJ* 2014;348:g3617-3617. doi:10.1136/bmj.g3617.

17 Manuel DG, Lim J, Tanuseputro P, et al. Revisiting Rose: strategies for reducing coronary heart disease. *BMJ* 2006;332:659-62. doi:10.1136/bmj.332.7542.659.

18 Scholes S, Bajekal M, Love H, et al. Persistent socioeconomic inequalities in cardiovascular risk factors in England over 1994-2008: a time-trend analysis of repeated cross-sectional data. *BMC Public Health* 2012;12:129. doi:10.1186/1471-2458-12-129.

19 White M, Adams J, Heywood P. How and why do interventions that increase health overall widen inequalities within populations? In: Babones SJ, ed. *Social inequality and public health*. Policy Press, 2009: 64-81, http://policypress.universitypressscholarship.com/view/10.1332/policypress/9781847423207.001.0001/upso-9781847423207-chapter-5doi:10.1332/policypress/9781847423207.003.0005.

20 Wallach-Kildemoes H, Diderichsen F, Krasnik A, Lange T, Andersen M. Is the high-risk strategy to prevent cardiovascular disease equitable? A pharmacoepidemiological cohort study. *BMC Public Health* 2012;12:610. doi:10.1186/1471-2458-12-610.

21 Bender AM, Kawachi I, Jørgensen T, Pisinger C. Neighborhood deprivation is strongly associated with participation in a population-based health check. *PLoS One* 2015;10:e0129819. doi:10.1371/journal.pone.0129819.

22 Chang KC-M, Soljak M, Lee JT, et al. Coverage of a national cardiovascular risk assessment and management programme (NHS Health Check): Retrospective database study. *Prev Med* 2015;78:1-8. doi:10.1016/j.ypmed.2015.05.022.

23 Gidlow C, Ellis N, Randall J, et al. Method of invitation and geographical proximity as predictors of NHS Health Check uptake. *J Public Health* 2014;26. doi:10.1093/pubmed/fdu092.

24 Attwood S, Morton K, Sutton S. Exploring equity in uptake of the NHS Health Check and a nested physical activity intervention trial. *J Public Health (Oxf)* 2015;01:fdv070. doi:10.1093/pubmed/fdv070.

25 Baker C, Loughren EA, Crone D, Kallfa N. A process evaluation of the NHS Health Check care pathway in a primary care setting. *J Public Health (Oxf)* 2015;37:202-9. doi:10.1093/pubmed/fdv053.

26 Unal B, Critchley JA, Capewell S. Explaining the decline in coronary heart disease mortality in England and Wales between 1981 and 2000. *Circulation* 2004;109:1101-7. doi:10.1161/01.CIR.0000118498.35499.B2.

27 Bajekal M, Scholes S, Love H, et al. Analysing recent socioeconomic trends in coronary heart disease mortality in England, 2000-2007: a population modelling study. *PLoS Med* 2012;9:e1001237. doi:10.1371/journal.pmed.1001237.

28 Ainsworth JD, Carruthers E, Couch P, et al. IMPACT: a generic tool for modelling and simulating public health policy. *Methods Inf Med* 2011;50:454-63. doi:10.3414/ME11-02-0006.

29 Department for Communities and Local Government. English indices of deprivation 2010 - Publications - GOV.UK. 2011. www.gov.uk/government/statistics/english-indices-of-deprivation-2010.

30 Mindell J, Biddulph JP, Hirani V, et al. Cohort profile: the health survey for England. *Int J Epidemiol* 2012;41:1585-93. doi:10.1093/ije/dyr199.

31 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2001*. [computer file] 3rd ed. UK Data Archive, 2010. [distributor], 10.5255/UKDA-SN-4628-1.

32 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2002*. [computer file] 2nd ed. UK Data Archive, 2010. [distributor], 10.5255/UKDA-SN-4912-1.

33 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2003*. [computer file] 2nd ed. UK Data Archive, 2010. [distributor], 10.5255/UKDA-SN-5098-1.

34 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2004*. [computer file] 2nd ed. UK Data Archive, 2010. [distributor], 10.5255/UKDA-SN-5439-1.

35 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2005*. [computer file] 3rd ed. UK Data Archive, 2011. [distributor], 10.5255/UKDA-SN-5675-1.

36 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2006*. [computer file] 4th ed. UK Data Archive, 2011. [distributor], 10.5255/UKDA-SN-5809-1.

37 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2007*. [computer file] 2nd ed. UK Data Archive, 2010. [distributor], 10.5255/UKDA-SN-6112-1.

38 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2008*. [computer file] 4th ed. UK Data Archive, 2013. [distributor], 10.5255/UKDA-SN-6397-2.

39 National Centre for Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2009*. [computer file] 2nd ed. UK Data Archive, 2011. [distributor], 10.5255/UKDA-SN-6732-1.

40 NatCen Social Research, Free Royal, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2010*. [computer file] 2nd ed. UK Data Archive, 2012. [distributor], 10.5255/UKDA-SN-6986-2.

41 NatCen Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2011*. [computer file] UK Data Archive, 2013. [distributor], 10.5255/UKDA-SN-7260-1.

42 NatCen Social Research, University College London. *Department of Epidemiology and Public Health. Health Survey for England, 2012*. [computer file] UK Data Archive, 2014. [distributor], 10.5255/UKDA-SN-7480-1.

43 Public Health England. Explore NHS Health Check Data. www.healthcheck.nhs.uk/interactive_map/.

44 Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. *BMJ* 2012;344:e4181. doi:10.1136/bmj.e4181.

45 Forster AS, Dodhia H, Booth H, et al. Estimating the yield of NHS Health Checks in England: a population-based cohort study. *J Public Health (Oxf)* 2015;37:234-40. doi:10.1093/pubmed/fdu079.

46 Wallach-Kildemoes H, Andersen M, Diderichsen F, Lange T. Adherence to preventive statin therapy according to socioeconomic position. *Eur J Clin Pharmacol* 2013;69:1553-63. doi:10.1007/s00228-013-1488-6.

47 Hutfless S, Gudzune KA, Maruthur N, et al. Strategies to prevent weight gain in adults: a systematic review. *Am J Prev Med* 2013;45:e41-51. doi:10.1016/j.amepre.2013.07.013.

48 Bauld L, Bell K, McCullough L, Richardson L, Greaves L. The effectiveness of NHS smoking cessation services: a systematic review. *J Public Health (Oxf)* 2010;32:71-82. doi:10.1093/pubmed/fdp074.

49 West R. Lesley Owen. Estimates of 52-week continuous abstinence rates following selected smoking cessation interventions in England. Smoking in England 2012. www.smokinginengland.info/downloadfile/?type=report&src=5.

50 Sharma A, Hauck K, Hollingsworth B, Siciliani L. The effects of taxing sugar-sweetened beverages across different income groups. *Health Econ* 2014;23:1159-84. doi:10.1002/hec.3070.

51 Briggs ADM, Mytton OT, Kehlbacher A, Tiffin R, Rayner M, Scarborough P. Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in UK: econometric and comparative risk assessment modelling study. *BMJ* 2013;347:f6189. doi:10.1136/bmj.f6189.

52 Cabrera Escobar MA, Veerman JL, Tollman SM, Bertram MY, Hofman KJ. Evidence that a tax on sugar sweetened beverages reduces the obesity rate: a meta-analysis. *BMC Public Health* 2013;13:1072. doi:10.1186/1471-2458-13-1072.

53 Webster JL, Dunford EK, Hawkes C, Neal BC. Salt reduction initiatives around the world. *J Hypertens* 2011;29:1043-50. doi:10.1097/HJH.0b013e328345ed83.

54 Gillespie DOS, Allen K, Guzman-Castillo M, et al. The health equity and effectiveness of policy options to reduce dietary salt intake in England: policy forecast. *PLoS One* 2015;10:e0127927. doi:10.1371/journal.pone.0127927.

55 Bartlett S, Klerman J, Olsha L, et al. Evaluation of the Healthy Incentives Pilot (HIP) Final Report. 2014. www.fns.usda.gov/sites/default/files/HIP-Final.pdf.

56 Nnoaham KE, Sacks G, Rayner M, Mytton O, Gray A. Modelling income group differences in the health and economic impacts of targeted food taxes and subsidies. *Int J Epidemiol* 2009;38:1324-33. doi:10.1093/ije/dyp214.

57 Levy DT, Huang A-T, Currie LM, Clancy L. The benefits from complying with the framework convention on tobacco control: a SimSmoke analysis of 15 European nations. *Health Policy Plan* 2014;29:1031-42. doi:10.1093/heapol/czt085.

58 Alfons A, Kraft S, Templ M, et al. Simulation of close-to-reality population data for household surveys with application to EU-SILC. *Stat Methods Appl* 2011;20:383-407. doi:10.1007/s10260-011-0163-2.

59 Mackenbach JP, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Soc Sci Med* 1997;44:757-71. doi:10.1016/S0277-9536(96)00073-1.

60 Office for National Statistics. Results of the ICD-10 v2010 bridge coding study, England and Wales. 2011. www.ons.gov.uk/ons/rel/subnational-health3/results-of-the-icd-10-v2010-bridge-coding-study--england-and-wales--2009/2009/index.html.

61 Schmid VJ, Held L. Bayesian age-period-cohort modeling and prediction-BAMP. *J Stat Softw* 2007;21:1-15doi:10.18637/jss.v021.i08.

62 Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. *J Epidemiol Community Health* 2013;67:190-3. doi:10.1136/jech-2012-201257.

63 McLaren L, McIntyre L, Kirkpatrick S. Rose's population strategy of prevention need not increase social inequalities in health. *Int J Epidemiol* 2010;39:372-7. doi:10.1093/ije/dyp315.

64 Scholes S, Bajekal M, Norman P, et al. Quantifying policy options for reducing future coronary heart disease mortality in England: a modelling study. *PLoS One* 2013;8:e69935. doi:10.1371/journal.pone.0069935.

65 Schuetz CA, Alperin P, Guda S, et al. A standardized vascular disease health check in europe: a cost-effectiveness analysis. *PLoS One* 2013;8:e66454. doi:10.1371/journal.pone.0066454.

66 Bandosz P, O'Flaherty M, Drygas W, et al. Decline in mortality from coronary heart disease in Poland after socioeconomic transformation: modelling study. *BMJ* 2012;344:d8136. doi:10.1136/bmj.d8136.

67 Franco M, Bilal U, Orduñez P, et al. Population-wide weight loss and regain in relation to diabetes burden and cardiovascular mortality in Cuba 1980-2010: repeated cross sectional surveys and ecological comparison of secular trends. *BMJ* 2013;346:f1515. doi:10.1136/bmj.f1515.

68 Danaei G, Finucane MM, Lin JK, et al. Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Blood Pressure). National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5·4 million participants. *Lancet* 2011;377:568-77. doi:10.1016/S0140-6736(10)62036-3.

69 Farzadfar F, Finucane MM, Danaei G, et al. Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Cholesterol). National, regional, and global trends in serum total cholesterol since 1980: systematic analysis of health examination surveys and epidemiological studies with 321 country-years and 3·0 million participants. *Lancet* 2011;377:578-86. doi:10.1016/S0140-6736(10)62038-7.

70 Molarius A, Parsons RW, Dobson AJ, et al. WHO MONICA Project. Trends in cigarette smoking in 36 populations from the early 1980s to the mid-1990s: findings from the WHO MONICA Project. *Am J Public Health* 2001;91:206-12. doi:10.2105/AJPH.91.2.206.

71 Maguire ER, Monsivais P. Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data. *Br J Nutr* 2015;113:181-9. doi:10.1017/S0007114514002621.

72 Barton P, Andronis L, Briggs A, McPherson K, Capewell S. Effectiveness and cost effectiveness of cardiovascular disease prevention in whole populations: modelling study. *BMJ* 2011;343:d4044. doi:10.1136/bmj.d4044.

73 Vos T, Carter R, Barendregt J, et al. *Assessing Cost-Effectiveness in Prevention.* The University of Queensland, Brisbane, and Deakin University, 2010, www.sph.uq.edu.au/docs/BODCE/ACE-P/ACE-Prevention_final_report.pdf.

74 Lloyd-Williams F, Bromley H, Orton L, et al. Smorgasbord or symphony? Assessing public health nutrition policies across 30 European countries using a novel framework. *BMC Public Health* 2014;14:1195. doi:10.1186/1471-2458-14-1195.

75 Bødker M, Pisinger C, Toft U, Jørgensen T. The rise and fall of the world's first fat tax. *Health Policy* 2015;119:737-42. doi:10.1016/j.healthpol.2015.03.003.

76 Leifert RM, Lucinda CR. Linear Symmetric 'Fat Taxes': Evidence from Brazil. *Appl Econ Perspect Policy* 2015;37:634-66. doi:10.1093/aepp/ppu062.

77 Madden D. The poverty effects of a 'fat-tax' in Ireland. *Health Econ* 2015;24:104-21. doi:10.1002/hec.3006.

78 The Marmot Review. *Fair society, healthy lives: strategic review of health inequalities in England post 2010.* UCL Institute of Health Equity, 2010.

79 Farr Institute. Investment in supercomputer accelerates the pace of health research across the North of England. www.farrinstitute.org/news/investment-in-supercomputer-accelerates-the-pace-of-health-research-across-the-north-of-england.

80 World Health Organization. Comparative quantification of health risks. WHO, 2004. www.who.int/healthinfo/global_burden_disease/cra/en/.

Appendix: supplementary information

REFERENCES

1. Global atlas on cardiovascular disease prevention and control. Geneva: World Health Organisation in collaboration with the World Heart Federation and the World Stroke Organisation, 2011. Available at: http://apps.who.int/iris/bitstream/10665/44701/1/9789241564373_eng.pdf (accessed on 24th July 2016) (cit. on p. 7).

2. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;385:117–171. DOI: 10.1016/S0140-6736(14)61682-2. (Cit. on pp. 7, 13, 15).

3. Roth GA, Forouzanfar MH, Moran AE et al. Demographic and epidemiologic drivers of global cardiovascular mortality. 2015;372:1333–1341. DOI: 10.1056/NEJMoa1406656 (cit. on pp. 7, 10).

4. Townsend N, Williams J, Bhatnagar P, Wickramasinghe K and Rayner M. Cardiovascular disease statistics 2014. London: British Heart Foundation, 2014. Available at: https://www.bhf.org.uk/~/media/files/publications/research/bhf_cvd-statistics-2014_web_2.pdf (accessed on 23rd June 2015) (cit. on p. 8).

5. Townsend N, Bhatnagar P, Wilkins E, Wickramasinghe K and Rayner M. Cardiovascular disease statistics 2015. London: British Heart Foundation, 2015. Available at: https://www.bhf.org.uk/publications/statistics/cvd-stats-2015 (cit. on pp. 8–12).

6. Smolina K, Wright FL, Rayner M and Goldacre MJ. Determinants of the decline in mortality from acute myocardial infarction in England between 2002 and 2010: linked national database study. BMJ 2012;344:d8059. DOI: 10.1136/bmj.d8059. (Cit. on pp. 8–11).

7. Smolina K, Wright FL, Rayner M and Goldacre MJ. Incidence and 30-day case fatality for acute myocardial infarction in England in 2010: national-linked database study. Eur J Public Health 2012;22:848–853. DOI: 10.1093/eurpub/ckr196. (Cit. on pp. 8, 58, 59).

8. Rothwell PM, Coull AJ, Silver LE et al. Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). Lancet 2005;366:1773–1783. DOI: 10.1016/S0140-6736(05)67702-1. (Cit. on p. 8).

9. Scarborough P, Peto V, Bhatnagar P et al. Stroke statistics 2009 edition. Research rep. British Heart Foundation, 2009. Available at: https://www.bhf.org.uk/~/media/files/publications/research/hs2009_stroke_statistics.pdf (accessed on 24th Mar. 2016) (cit. on p. 8).

10. Pace M, Cayotte E, Agafitei L et al. Revision of the European Standard Population report of Eurostat's task force: 2013 edition. Luxembourg: Eurostat, 2013. DOI: 10.2785/11470. (Cit. on pp. 9, 10, 14, 16, 25).

11. Boniface S, Bridges S, Craig R et al. Health Survey for England 2011. Health, social care and lifestyles. Joint Health Surveys Unit NatCen Social Research Department of Epidemiology and Public Health, UCL (University College London) Leeds, UK. Health and Social Care Information Centre, 2012. Available at: http://www.hscic.gov.uk/catalogue/PUB09300 (cit. on pp. 9, 24).

12. Unal B, Critchley JA and Capewell S. Modelling the decline in coronary heart disease deaths in England and Wales, 1981-2000: comparing contributions from primary prevention and secondary prevention. BMJ 2005;331:614. DOI: 10.1136/bmj.38561.633345.8F. (Cit. on p. 10).

13. Lee S, Shafe ACE and Cowie MR. UK stroke incidence, mortality and cardiovascular risk management 1999–2008: time-trend analysis from the General Practice Research Database. BMJ Open 2011;1:e000269. DOI: 10.1136/bmjopen-2011-000269. (Cit. on pp. 11, 59).

14. World Health Organisation. Cancer fact sheet No297. 2015. Available at: http://www.who.int/mediacentre/factsheets/fs297/en/ (accessed on 24th Mar. 2016) (cit. on p. 13).

15. Office for National Statistics. Deaths registered in England and Wales. 2014. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsand marriages/deaths/bulletins/deathsregistrationsummarytables/2015-07-15 (accessed on 24th Mar. 2016) (cit. on p. 13).

16. Cancer Research UK. Cancer statistics for the UK. 2016. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics (accessed on 4th Feb. 2016) (cit. on pp. 13, 14).

17. Office for National Statistics. Cancer registration statistics quality and methodology information. Cancer registration statistics quality and methodology information. 2016. Available at: http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/qmis/cancerregistrationstatisticsqmi (accessed on 25th Mar. 2016) (cit. on p. 13).

18. Ferlay J, Soerjomataram I, Dikshit R et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2014;136:E359–E386. DOI: 10.1002/ijc.29210 (cit. on pp. 13, 15).

19. Cancer Research UK. Lung cancer statistics. 2016. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer (accessed on 11th July 2016) (cit. on pp. 14, 15).

20. Newton JN, Briggs ADM, Murray CJL et al. Changes in health in England, with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;386:2257–2274. DOI: 10.1016/S0140-6736(15)00195-6. (Cit. on pp. 14, 17, 23, 36, 147).

21. Allemani C, Weir HK, Carreira H et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet 2015;385:977–1010. DOI: 10.1016/S0140-6736(14)62038-9. (Cit. on pp. 15, 17, 60, 137, 150).

22. Maddams J, Brewster D, Gavin A et al. Cancer prevalence in the United Kingdom: estimates for 2008. Br J Cancer 2009;101:541–547. DOI: 10.1038/sj.bjc.6605148. (Cit. on p. 15).

23. British Lung Foundation. Lung cancer statistics. 2016. Available at: https://statistics.blf.org.uk/lung-cancer (accessed on 12th July 2016) (cit. on p. 15).

24. Continuous Update Project report: diet, nutrition, physical activity and stomach cancer. Research rep. World Cancer Research Fund International/American Institute for Cancer Research, 2016. Available at: http://wcrf.org/sites/default/files/Stomach-Cancer-2016-Report.pdf (accessed on 22nd Apr. 2016) (cit. on pp. 15, 19–22, 73, 75, 193).

25. Cancer Research UK. Stomach cancer statistics. 2016. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/stomach-cancer (accessed on 11th July 2016) (cit. on pp. 16, 17).

26. Coupland VH, Allum W, Blazeby JM et al. Incidence and survival of oesophageal and gastric cancer in England between 1998 and 2007, a population-based study. BMC Cancer 2012;12:11. DOI: 10.1186/1471-2407-12-11. (Cit. on p. 16).

27. Forman D, Stockton D, Møller H et al. Cancer prevalence in the UK: results from the EUROPREVAL study. Ann Oncol 2003;14:648–654. DOI: 10.1093/annonc/mdg169. (Cit. on p. 17).

28. National Cancer Intelligence Network. Cancer prevalence - one, five and ten year data. Available at: http://www.ncin.org.uk/Prevalence/1_5_10_Year/atlas.html (accessed on 14th July 2016) (cit. on p. 17).

29. Alwan A, Armstrong T, Bettcher D et al. Global status report on noncommunicable diseases 2010. Geneva: World Health Organisation, 2011. Available at: http://whqlibdoc.who.int/publications/2011/9789240686458_eng.pdf (cit. on pp. 17, 135).

30. Comparative quantification of health risks. Geneva: World Health Organisation, 2004. Available at: http://www.who.int/publications/cra/en/ (accessed on 30th Jan. 2014) (cit. on pp. 17, 36, 49, 67, 134, 174).

31. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington, DC: World Cancer Research Fund and American Institute for Cancer Research, 2007 (cit. on pp. 17, 19, 21, 70, 194).

32. Danaei G, Ding EL, Mozaffarian D et al. The preventablecauses of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLOS Med 2009;6. Ed. by Hales S:e1000058. DOI: 10.1371/journal.pmed.1000058. (Cit. on pp. 17, 36, 67, 174).

33. Forouzanfar MH, Alexander L, Anderson HR et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;386:2287–2323. DOI: 10.1016/S0140-6736(15)00128-2. (Cit. on pp. 17, 36, 147).

34. Roffo AH. Durch tabak beim kaninchen entwickeltes carcinom [Tobacco-induced cancer in rabbits]. Zeitschrift für Krebsforschung 1931;33:321 (cit. on p. 18).

35. Roffo AH. Krebserzeugende tabakwirkung [The carcinogenic effects of tobacco]. Monatsschrift für Krebsbekämpfung 1940;8:97–102. Available at: http://www.scielosp.org/pdf/bwho/v84n6/v84n6a21.pdf (cit. on p. 18).

36. Müller FH. Tabakmissbrauch und lungencarcinom [Tobacco misuse and lung carcinoma]. Zeitschrift für Krebsforschung 1939;49:5785 (cit. on p. 18).

37. Ezzati M, Henley SJ, Thun MJ and Lopez AD. Role of smoking in global and regional cardiovascular mortality. Circulation 2005;112:489–497. DOI: 10.1161/CIRCULATIONAHA.104.521708. (Cit. on pp. 18, 49, 71, 187, 190).

38. Ezzati M and Lopez AD. Comparative quantification of health risks. Chapter 11: smoking and oral tobacco use. Geneva: World Health Organisation, 2004. Available at: http://www.who.int/publications/cra/en/ (cit. on p. 18).

39. International Agency for Research on Cancer working group on the evaluation of carcinogenic risks to humans. Tobacco smoke and involuntary smoking. Lyon, France; Geneva: International Agency for Research on Cancer and World Health Organisation, 2004. Available at: http://monographs.iarc.fr/ENG/Monographs/vol83/mono83.pdf (cit. on p. 18).

40. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The health consequences of smoking–50 years of progress: a report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US), 2014. Available at: http://www.surgeongeneral.gov/library/reports/50-years-of-progress/full-report.pdf (accessed on 16th July 2016) (cit. on p. 18).

41. Nabi H, Estaqiuo C and Auleley GR. Smoking and mortality–beyond established causes. N Engl J Med 2015;372:2169. DOI: 10.1056/NEJMc1503675\\#SA3 (cit. on pp. 18, 147).

42. Wolf PA, D'Agostino RB, Kannel WB, Bonita R and Belanger AJ. Cigarette smoking as a risk factor for stroke: the Framingham study. JAMA 1988;259:1025–1029. DOI: 10.1001/jama.1988.03720070025028. (Cit. on pp. 18, 72).

43. Health and Social Care Information Centre. Statistics on smoking: England, 2016. Health and Social Care Information Centre, 2016. Available at: http://www.hscic.gov.uk/catalogue/PUB20781/stat-smok-eng-2016-rep.pdf (accessed on 16th July 2016) (cit. on pp. 18, 24, 147).

44. Health and Social Care Information Centre. Health Survey for England, 2014: trend tables. 2015. Available at: http://www.hscic.gov.uk/catalogue/PUB19297 (accessed on 17th July 2016) (cit. on pp. 18, 23).

45. Schoenfeld JD and Ioannidis JPA. Is everything we eat associated with cancer? A systematic cookbook review. Am J Clin Nutr 2013;97:127–134. DOI: 10.3945/ajcn.112.047142. (Cit. on p. 19).

46. Kromhout D, Spaaij CJK, de Goede J and Weggemans RM. The 2015 Dutch food-based dietary guidelines. Eur J Clin Nutr 2016;70:869–878. DOI: 10.1038/ejcn.2016.52. (Cit. on p. 19).

47. Mozaffarian D. Dietary and policy priorities for cardiovascular disease, diabetes, and obesity a comprehensive review. Circulation 2016;133:187–225. DOI: 10.1161/circulationaha.115.018585. (Cit. on p. 19).

48. Lock K, Pomerleau J, Causer L and McKee M. Comparative quantification of health risks. Chapter 9: low fruit and vegetable consumption. Geneva: World Health Organisation, 2004. Available at: http://www.who.int/publications/cra/en/ (cit. on pp. 19, 75, 193).

49. Wang X, Ouyang Y, Liu J et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 2014;349:g4490. DOI: 10.1136/bmj.g4490. (Cit. on pp. 19, 74).

50. Oyebode O, Gordon-Dseagu V, Walker A and Mindell JS. Fruit and vegetable consumption and all-cause, cancer and CVD mortality: analysis of Health Survey for England data. J Epidemiol Community Health 2014;68:856–862. DOI: 10.1136/jech-2013-203500 (cit. on pp. 19, 74).

51. Kypridemos C, O'Flaherty M and Capewell S. Fruit and vegetable consumption and non-communicable disease: time to update the '5 a day' message? J Epidemiol Community Health 2014;68:799–800. DOI: 10.1136/jech-2014-203981 (cit. on p. 19).

52. Vieira AR, Abar L, Vingeliene S et al. Fruits, vegetables and lung cancer risk: a systematic review and meta-analysis. Ann Oncol 2016;27:81–96. DOI: 10.1093/annonc/mdv381. (Cit. on pp. 19, 74, 193).

53. Wang Y, Li F, Wang Z, Qiu T, Shen Y and Wang M. Fruit and vegetable consumption and risk of lung cancer: a dose–response meta-analysis of prospective cohort studies. Lung Cancer 2015;88:124–130. DOI: 10.1016/j.lungcan.2015.02.015. (Cit. on pp. 19, 74).

54. Public Health England and Food Standards Agency. NDNS results from years 1 to 4 combined of the rolling programme for 2008 and 2009 to 2011 and 2012: report. London: Public Health England and the Food Standards Agency, 2014. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/310995/NDNS_Y1_to_4_UK_report.pdf (accessed on 17th July 2016) (cit. on pp. 20, 174).

55. Roberts C. Health Survey for England 2013 - Health, social care and lifestyles. Chapter 7: fruit and vegetable consumption. Joint Health Surveys Unit NatCen Social Research Department of Epidemiology and Public Health, UCL (University College London) Leeds, UK. London: Health and Social Care Information Centre, 2014. Available at: http://www.hscic.gov.uk/catalogue/PUB16076/HSE2013-Ch7-fru-veg-com.pdf (cit. on p. 20).

56. Strazzullo P, D'Elia L, Kandala NB and Cappuccio FP. Salt intake, stroke, and cardiovascular disease: meta-analysis of prospective studies. BMJ 2009;339:b4567. DOI: 10.1136/bmj.b4567. (Cit. on pp. 20, 135).

57. D'Elia L, Rossi G, Ippolito R, Cappuccio FP and Strazzullo P. Habitual salt intake and risk of gastric cancer: a meta-analysis of prospective studies. Clin Nutr 2012;31:489–498. DOI: 10.1016/j.clnu.2012.01.003. (Cit. on pp. 20, 135).

58. He FJ, Li J and MacGregor GA. Effect of longer-term modest salt reduction on blood pressure. In: *Cochrane Database Syst Rev*. John Wiley & Sons, Ltd, 2013. DOI: 10.1002/14651858.cd004937.pub2. (Cit. on p. 20).

59. Mozaffarian D, Fahimi S, Singh GM et al. Global sodium consumption and death from cardiovascular causes. N Engl J Med 2014;371:624–634. DOI: 10.1056/NEJMoa1304127. (Cit. on pp. 20, 70, 135, 137, 196).

60. Ioannidis JPA. Commentary: salt and the assault of opinion on evidence. Int J Epidemiol 2016;45:264–265. DOI: 10.1093/ije/dyw015. (Cit. on pp. 20, 145).

61. Guideline: sodium intake for adults and children. Geneva: World Health Organisation, 2012. Available at: http://www.ncbi.nlm.nih.gov/books/NBK133309/ (accessed on 17th July 2016) (cit. on p. 20).

62. Scientific Advisory Committee on Nutrition. Salt and health. Norwich, UK: The stationery office, 2003. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/338782/SACN_Salt_and_Health_report.pdf (cit. on pp. 20, 135, 144).

63. Mente A, O'Donnell M, Rangarajan S et al. Associations of urinary sodium excretion with cardiovascular events in individuals with and without hypertension: a pooled analysis of data from four studies. Lancet 2016;388:465–475. DOI: 10.1016/s0140-6736(16)30467-6 (cit. on pp. 20, 145).

64. O'Donnell M, Mente A, Rangarajan S et al. Urinary sodium and potassium excretion, mortality, and cardiovascular events. N Engl J Med 2014;371:612–623. DOI: 10.1056/NEJMoa1311889. (Cit. on pp. 20, 145).

65. Cogswell ME, Mugavero K, Bowman BA and Frieden TR. Dietary sodium and cardiovascular disease risk — measurement matters. N Engl J Med 2016;375:580–586. DOI: 10.1056/NEJMsb1607161. (Cit. on pp. 20, 145).

66. Taylor RS, Ashton KE, Moxham T, Hooper L and Ebrahim S. Reduced dietary salt for the prevention of cardiovascular disease: a meta-analysis of randomized controlled trials (Cochrane review). Am J Hypertens 2011;24:843–853. DOI: 10.1038/ajh.2011.115 (cit. on p. 20).

67. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP and Meerpohl JJ. Effect of lower sodium intake on health: systematic review and meta-analyses. BMJ 2013;346:f1326. DOI: 10.1136/bmj.f1326. (Cit. on p. 20).

68. Sadler K, Nicholson S, Steer T et al. National Diet and Nutrition Survey - Assessment of dietary sodium in adults (aged 19 to 64 years) in England, 2011. London: Department of Health, 2012. Available at: http://webarchive.nationalarchives.gov.uk/20130402145952/http://transparency.dh.gov.uk/2012/06/21/sodium-levels-among-adults/ (cit. on pp. 21, 51, 135, 137, 226, 227).

69. Bull FC, Armstrong TP, Dixon T, Ham S, Neiman A and Pratt M. Comparative quantification of health risks. Chapter 10: physical inactivity. Geneva: World Health Organisation, 2004. Available at: http://www.who.int/publications/cra/en/ (cit. on pp. 21, 74, 190, 192).

70. Arem H, Moore SC, Patel A and al et. Leisure time physical activity and mortality: A detailed pooled analysis of the dose-response relationship. JAMA Internal Medicine 2015;175:959. DOI: 10.1001/jamainternmed.2015.0533. (Cit. on p. 21).

71. Fiona Bull, Stuart Biddle, David Buchner et al. Physical activity guidlines in the UK: review and recommendations. London: Department of Health, 2010. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213743/dh_128255.pdf (accessed on 18th July 2016) (cit. on p. 21).

72. Scholes S and Mindell J. Health Survey for England 2012. Chapter 2: physical activity in adults. Health and Social Care Information Centre, 2012. Available at: http://www.hscic.gov.uk/catalogue/PUB13218/HSE2012-Ch2-Phys-act-adults.pdf (accessed on 18th July 2016) (cit. on p. 21).

73. James WPT, Jackson-Leach R, Mhurchu CN et al. Comparative quantification of health risks. Chapter 8: overweight and obesity (high body mass index). Geneva: World Health Organisation, 2004. Available at: http://www.who.int/publications/cra/en/ (cit. on pp. 21, 22).

74. The Emerging Risk Factors Collaboration. Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. Lancet 2011;377:1085–1095. DOI: 10.1016/S0140-6736(11)60105-0. (Cit. on pp. 21, 73, 175, 189, 192).

75. Smith L, Brinton LA, Spitz MR et al. Body mass index and risk of lung cancer among never, former, and current smokers. J Natl Cancer Inst 2012;104:778–789. DOI: 10.1093/jnci/djs179 (cit. on p. 22).

76. Tammemägi MC, Katki HA, Hocking WG et al. Selection criteria for lung-cancer screening. N Engl J Med 2013;368:728–736. DOI: 10.1056/NEJMoa1211776. (Cit. on pp. 22, 71, 72).

77.   Scantlebury R and Moody A. Health Survey for England 2012. Chapter 9: adult obesity and overweight. Health and Social Care Information Centre, 2015. Available at: http://www.hscic.gov.uk/catalogue/PUB13218/HSE2012-Ch2-Phys-act-adults.pdf (accessed on 18th July 2016) (cit. on p. 22).

78.   Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002;360:1903–1913. DOI: 10.1016/S0140-6736(02)11911-8. (Cit. on pp. 22, 73, 175, 188, 191).

79.   Lawes CM, Hoorn SV, Law MR, Elliott P, MacMahon S and Rodgers A. Comparative quantification of health risks. Chapter 6: high blood pressure. Geneva: World Health Organisation, 2004. Available at: http://www.who.int/publications/cra/en/ (cit. on p. 22).

80.   Health and Social Care Information Centre. Health Survey for England - 2006: latest trends. 2008. Available at: http://www.hscic.gov.uk/pubs/hse06trends (accessed on 20th July 2016) (cit. on p. 23).

81.   Prospective Studies Collaboration. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55 000 vascular deaths. Lancet 2007;370:1829–1839. DOI: 10.1016/S0140-6736(07)61778-4. (Cit. on pp. 23, 73, 175, 189, 192).

82.   Lawes CM, Hoorn SV, Law MR and Rodgers A. Comparative quantification of health risks. Chapter 7: high cholesterol. Geneva: World Health Organisation, 2004. Available at: http://www.who.int/publications/cra/en/ (cit. on p. 23).

83.   Oyinlola Oyebode. Health Survey for England 2011. Chapter 2: CVD. Joint Health Surveys Unit NatCen Social Research Department of Epidemiology and Public Health, UCL (University College London) Leeds, UK. Health and Social Care Information Centre, 2012. Available at: http://www.hscic.gov.uk/catalogue/PUB09300/HSE2011-Ch2-CVD.pdf (cit. on pp. 23, 24, 103).

84.   The Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group. Global burden of metabolic risk factors of chronic diseases. 2014. Available at: http://www1.imperial.ac.uk/publichealth/departments/ebs/projects/eresh/majidezzati/healthmetrics/metabolicriskfactors/ (accessed on 6th Oct. 2014) (cit. on p. 23).

85.   Kawachi I, Subramanian SV and Almeida-Filho N. A glossary for health inequalities. J Epidemiol Community Health 2002;56:647–652. DOI: 10.1136/jech.56.9.647. (Cit. on pp. 23, 24).

86.   Bennett JE, Li G, Foreman K et al. The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. Lancet 2015;386:163–170. DOI: 10.1016/S0140-6736(15)60296-3 (cit. on p. 23).

87. Dahlgren G and Whitehead M. European strategies for tackling social inequities in health: levelling up part 2. Research rep. Copenhagen: World Health Organisation, 2006. Available at: http://www.hcsunderland.org.uk/content/ESFTSIH%20PART2.pdf (accessed on 11th Feb. 2015) (cit. on pp. 24, 26).

88. Roberts K, Cavill N, Hancock C and Rutter H. Social and economic inequalities in diet and physical activity. PHE publications gateway number: 201 3 257. London, UK: Public Health England, 2013. Available at: http://www.noo.org.uk/ (cit. on pp. 24, 103, 132).

89. Maguire ER and Monsivais P. Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data. Br J Nutr 2015;113:181–189. DOI: 10.1017/S0007114514002621. (Cit. on pp. 24, 132).

90. Ji C and Cappuccio FP. Socioeconomic inequality in salt intake in Britain 10 years after a national salt reduction programme. BMJ Open 2014;4:e005683. DOI: 10.1136/bmjopen-2014-005683. (Cit. on pp. 24, 135, 144).

91. Ji C, Kandala NB and Cappuccio FP. Spatial variation of salt intake in Britain and association with socioeconomic status. BMJ Open 2013;3:e002246. DOI: 10.1136/bmj open-2012-002246. (Cit. on pp. 24, 135).

92. Townsend N, Wickramasinghe K, Bhatnagar P et al. Coronary heart disease statistics 2012 edition. London. British Heart Foundation, 2012. Available at: http://www.bhf.org.uk/publications/view-publication.aspx?ps=1002097 (cit. on pp. 24, 58–60, 124, 137, 150).

93. National Cancer Intelligence Network and Cancer Research UK. Cancer by deprivation in England 1996 - 2011. London: Public Health England, 2014. Available at: http://www.ncin.org.uk/about_ncin/cancer_by_deprivation_in_england (cit. on pp. 24, 137, 150).

94. Office for National Statistics. Standard occupational classification and socio-economic classification. 2012. Available at: http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/soc-and-sec-archive/index.html (accessed on 27th Aug. 2014) (cit. on pp. 25, 105).

95. White C, Edgar G and Siegler V. Social inequalities in male mortality for selected causes of death by the National Statistics Socio-economic Classification, England and Wales, 2001-03. Health Stat Q 2008:19–32. Available at: https://www.ons.gov.uk/ons/rel/hsq/health-statistics-quarterly/no--38--summer-2008/health-statistics-quarterly.pdf#page=19 (cit. on p. 25).

96. Langford A, Johnson B and Al-Hamad A. Social inequalities in female mortality by region and by selected causes of death, England and Wales, 2001–03. Health Stat Q 2009;44:7–26. DOI: 10.1057/hsq.2009.35. (Cit. on p. 25).

97. WHO Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health: Commission on Social Determinants of Health final report. Geneva, Switzerland: World Health Organisation, 2008 (cit. on pp. 24, 28).

98. Marmot M, Allen J, Goldblatt P et al. Fair society, healthy lives: strategic review of health inequalities in England post 2010. London: UCL Institute of Health Equity, 2010 (cit. on pp. 24, 32, 33, 133, 134).

99. Dahlgren G and Whitehead M. Policies and strategies to promote social equity in health. Copenhagen: World Health Organisation, 1992. Available at: http://core.ac.uk/download/pdf/6472456.pdf (accessed on 5th May 2015) (cit. on p. 26).

100. Diderichsen F, Evans T and Whitehead M. The social origins of disparities in health. In: *Challenging inequities in health: from ethics to action.* New York: Oxford University Press, 2001 (cit. on pp. 26, 27).

101. Lynch JW, Kaplan GA, Cohen RD, Tuomilehto J and Salonen JT. Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality, and acute myocardial infarction? Am J Epidemiol 1996;144:934–942. Available at: http://aje.oxfordjournals.org/content/144/10/934 (accessed on 23rd Apr. 2015) (cit. on p. 28).

102. Theodorson T. Cardiovascular risk and risk reduction: a review of recent literature. J Family Community Med 1995;2:19 (cit. on p. 28).

103. Berry JD, Dyer A, Cai X et al. Lifetime risks of cardiovascular disease. N Engl J Med 2012;366:321–329. DOI: 10.1056/NEJMoa1012848. (Cit. on pp. 28, 175).

104. Capewell S, Livingston BM, MacIntyre K et al. Trends in case-fatality in 117 718 patients admitted with acute myocardial infarction in Scotland. Eur Heart J 2000;21:1833–1840. DOI: 10.1053/euhj.2000.2318 (cit. on p. 28).

105. MacIntyre K, Capewell S, Stewart S et al. Evidence of improving prognosis in heart failure: trends in case fatality in 66 547 patients hospitalized between 1986 and 1995. Circulation 2000;102:1126–1131. DOI: 10.1161/01.cir.102.10.1126 (cit. on p. 28).

106. Capewell S, MacIntyre K, Stewart S et al. Age, sex, and social trends in out-of-hospital cardiac deaths in Scotland 1986-95: a retrospective cohort study. Lancet 2001;358:1213–1217. DOI: 10.1016/S0140-6736(01)06343-7 (cit. on p. 28).

107. Riaz SP, Horton M, Kang J, Mak V, Lüchtenborg M and Møller H. Lung cancer Incidence and survival in England: an analysis by socioeconomic deprivation and urbanization. J Thorac Oncol 2011;6:2005–2010. DOI: 10.1097/JTO.0b013e31822b02db. (Cit. on pp. 28, 60, 150).

108. Forrest LF, Adams J, Wareham H, Rubin G and White M. Socioeconomic inequalities in lung cancer treatment: systematic review and meta-analysis. PLOS Med 2013;10:1–25. DOI: 10.1371/journal.pmed.1001376. (Cit. on pp. 28, 60, 150).

109. Stephens MR, Blackshaw GRJC, Lewis WG et al. Influence of socio-economic deprivation on outcomes for patients diagnosed with gastric cancer. Scand J Gastroenterol 2005;40:1351–1357. DOI: 10.1080/00365520510023666 (cit. on pp. 28, 60, 137).

110. Choi IK, Hyun JJ, Kim SY et al. Influence of socioeconomic status on survival and clinical outcomes in patients with advanced gastric cancer after chemotherapy. Oncol Res Treat 2014;37:310–314. DOI: 10.1159/000362625 (cit. on pp. 28, 60).

111. McGill R, Anwar E, Orton L et al. Are interventions to promote healthy eating equally effective for all? Systematic review of socioeconomic inequalities in impact. BMC Public Health 2015;15:457. DOI: 10.1186/s12889-015-1781-7. (Cit. on pp. 29, 32, 33, 135).

112. Lorenc T, Petticrew M, Welch V and Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. J Epidemiol Community Health 2013;67:190–193. DOI: 10.1136/jech-2012-201257. (Cit. on pp. 29, 32, 33, 131, 135, 166).

113. Regidor E. Measures of health inequalities: part 2. J Epidemiol Community Health 2004;58:900–903. DOI: 10.1136/jech.2004.023036. (Cit. on p. 29).

114. Regidor E. Measures of health inequalities: part 1. J Epidemiol Community Health 2004;58:858–861. DOI: 10.1136/jech.2003.015347. (Cit. on p. 29).

115. Mackenbach JP and Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Soc Sci Med 1997;44:757–771. DOI: 10.1016/S0277-9536(96)00073-1. (Cit. on pp. 29, 65).

116. Harper S, King NB, Meersman SC, Reichman ME, Breen N and Lynch J. Implicit value judgments in the measurement of health inequalities. Milbank Q 2010;88:4–29. DOI: 10.1111/j.1468-0009.2010.00587.x. (Cit. on p. 29).

117. Asada Y. On the choice of absolute or relative inequality measures. Milbank Q 2010;88:616–622. DOI: 10.1111/j.1468-0009.2010.00614.x. (Cit. on p. 29).

118. Kjellsson G, Gerdtham UG and Petrie D. Lies, damned lies, and health inequality measurements. Epidemiology 2015;26:673–680. DOI: 10.1097/EDE.0000000000000319. (Cit. on p. 29).

119. Oliver A. Reflections on the development of health inequalities policy in England. Health Care Anal 2010;18:402–420. DOI: 10.1007/s10728-010-0144-x. (Cit. on p. 30).

120. Black D, Morris J, Smith C and Townsend P. Inequalities in health: report of a research rorking group chaired by Sir Douglas Black. Research rep. London: Inequalities in health: report of a research rorking group chaired by Sir Douglas Black, 1980 (cit. on p. 30).

121. Acheson D. Report of the Committee of Inquiry into the future development of the Public Health functions and Community Medicine. London: HMSO, 1988 (cit. on p. 30).

122. Department of Health. Department of Health equality objectives 2012 to 2016. 2012. Available at: https://www.gov.uk/government/publications/department-of-health-equality-objectives-2012-to-2016 (accessed on 6th Apr. 2016) (cit. on p. 30).

123. Improving outcomes: a strategy for cancer. Department of Health, 2011. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213785/dh_123394.pdf (accessed on 26th July 2016) (cit. on p. 30).

124. Department of Health. Cardiovascular disease outcomes strategy: improving outcomes for people with or at risk of cardiovascular disease. 2013. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/217118/9387-2900853-CVD-Outcomes_web1.pdf (accessed on 5th Dec. 2014) (cit. on pp. 30, 121, 131).

125. National Institute for Health and Care Excellence. Community engagement: improving health and wellbeing and reducing health inequalities [NG44]. 2016. Available at: https://www.nice.org.uk/guidance/ng44 (accessed on 26th July 2016) (cit. on p. 30).

126. National Institute for Health and Care Excellence. Cardiovascular disease prevention [PH25]. 2010. Available at: https://www.nice.org.uk/guidance/ph25 (accessed on 26th July 2016) (cit. on p. 30).

127. Marmot M, Allen J, Bell R and Goldblatt P. Building of the global movement for health equity: from Santiago to Rio and beyond. Lancet 2012;379:181–188. DOI: 10.1016/S0140-6736(11)61506-7. (Cit. on p. 30).

128. Rose G. Strategy of prevention: lessons from cardiovascular disease. BMJ 1981;282:1847–1851. DOI: 10.1136/bmj.282.6279.1847. (Cit. on p. 31).

129. Rose G. Sick individuals and sick populations. Int J Epidemiol 2001;30:427–432. DOI: 10.1093/ije/30.3.427. (Cit. on pp. 31, 131).

130. Rose G. The strategy of preventive medicine. 1992 (cit. on p. 31).

131. Barton P, Andronis L, Briggs A, McPherson K and Capewell S. Effectiveness and cost effectiveness of cardiovascular disease prevention in whole populations: modelling study. BMJ 2011;343:d4044. DOI: 10.1136/bmj.d4044. (Cit. on pp. 31, 133, 144, 166, 171).

132. Cooney MT, Dudina A, Whincup P et al. Re-evaluating the Rose approach: comparative benefits of the population and high-risk preventive strategies. Eur J Cardiovasc Prev Rehabil 2009;16:541–549. DOI: 10.1097/HJR.0b013e32832b38a1 (cit. on pp. 31, 166, 171).

133. Huffman MD, Lloyd-Jones DM, Ning H et al. Quantifying options for reducing coronary heart disease mortality by 2020. Circulation 2012;127:2477–2484. DOI: 10.1161/CIRCULATIONAHA.112.000769. (Cit. on pp. 31, 166, 171).

134. Pietinen P, Vartiainen E, Seppänen R, Aro A and Puska P. Changes in diet in Finland from 1972 to 1992: impact on coronary heart disease risk. Prev Med 1996;25:243–250. DOI: 10.1006/pmed.1996.0053. (Cit. on pp. 31, 103, 166, 171).

135. Puska P, Vartiainen E, Laatikainen T, Jousilahti P and Paavola M. The North Karelia project: from North Karelia to national action. [Helsinki]: National Institute for Health and Welfare, 2009 (cit. on pp. 31, 34, 166, 167, 171).

136. Emberson J, Whincup P, Morris R, Walker M and Ebrahim S. Evaluating the impact of population and high-risk strategies for the primary prevention of cardiovascular disease. Eur Heart J 2004;25:484–491. DOI: 10.1016/j.ehj.2003.11.012. (Cit. on pp. 31, 166, 171).

137. Manuel DG, Lim J, Tanuseputro P et al. Revisiting Rose: strategies for reducing coronary heart disease. BMJ 2006;332:659–662. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1403258/ (accessed on 6th Apr. 2015) (cit. on p. 31).

138. Manuel DG and Rosella LC. Commentary: assessing population (baseline) risk is a cornerstone of population health planning—looking forward to address new challenges. Int J Epidemiol 2010;39:380–382. DOI: 10.1093/ije/dyp373. (Cit. on p. 31).

139. Frohlich KL and Potvin L. The inequality paradox: the population approach and vulnerable populations. Am J Public Health 2008;98:216–221. DOI: 10.2105/AJPH.2007.114777. (Cit. on p. 31).

140. McLaren L, McIntyre L and Kirkpatrick S. Rose's population strategy of prevention need not increase social inequalities in health. Int J Epidemiol 2010;39:372–377. DOI: 10.1093/ije/dyp315. (Cit. on pp. 32, 131, 135).

141. Giddens A. The constitution of society: outline of the theory of structuration. University of California Press, 1986. 417 pp. Available at: http://www.ebook.de/de/product/3718417/anthony_giddens_anthony_goddens_the_constitution_of_society_outline_of_the_theory_of_structuration.html (cit. on p. 32).

142. Blankenship KM, Friedman SR, Dworkin S and Mantell JE. Structural interventions: concepts, challenges and opportunities for research. J Urban Health 2006;83:59–72. DOI: 10.1007/s11524-005-9007-4. (Cit. on p. 32).

143. Capewell S and Graham H. Will cardiovascular disease prevention widen health inequalities? PLOS Med 2010;7:e1000320. DOI: 10.1371/journal.pmed.1000320. (Cit. on pp. 32, 103, 121, 131, 132, 166).

144. White M, Adams J and Heywood P. How and why do interventions that increase health overall widen inequalities within populations? In: *Social Inequality and Public Health*. Ed. by Babones SJ. Policy Press, 2009:64–81. DOI: 10.1332/policypress/9781847423207.003.0005. (Cit. on pp. 32, 33, 132, 166).

145. Blaxter M. Evidence for the effect on inequalities in health of interventions designed to change behaviour. Bristol: Department of Social Medicine, University of Bristol, 2007. Available at: http://www.gserve.nice.org.uk/nicemedia/pdf/Evidencefor theeffectonInequalitiesdesignedtochangebehavior.pdf (accessed on 3rd June 2014) (cit. on pp. 32, 166).

146. Adams J, Mytton O, White M and Monsivais P. Why are some population interventions for diet and obesity more equitable and effective than others? The role of individual agency. PLOS Med 2016;13:e1001990. DOI: 10.1371/journal.pmed.1001990. (Cit. on pp. 32, 135, 166, 167).

147. Tugwell P, Savigny D de, Hawker G and Robinson V. Applying clinical epidemiological methods to health equity: the equity effectiveness loop. BMJ 2006;332:358–361. DOI: 10.1136/bmj.332.7537.358. (Cit. on pp. 32, 166).

148. Graham H. Tackling inequalities in health in England: remedying health disadvantages, narrowing health gaps or reducing health gradients? J Soc Policy 01 2004;33:115–131. DOI: 10.1017/S0047279403007220. (Cit. on p. 32).

149. Benach J, Malmusi D, Yasui Y and Martínez JM. A new typology of policies to tackle health inequalities and scenarios of impact based on Rose's population approach. J Epidemiol Community Health 2013;67:286–291. DOI: 10.1136/jech-2011-200363. (Cit. on p. 32).

150. Macintyre S. Inequalities in health in Scotland: what are they and what can we do about them. Research rep. Glasgow: MRC Social and Public Health Sciences Unit, 2007. Available at: http://www.sphsu.mrc.ac.uk/reports/OP017.pdf (cit. on pp. 33, 134).

151. Borkhoff CM, Saskin R, Rabeneck L et al. Disparities in receipt of screening tests for cancer, diabetes and high cholesterol in Ontario, Canada: a population-based study using area-based methods. Can J Public Health 2013;104:e284–290 (cit. on p. 33).

152. Honein-AbouHaidar GN, Baxter NN, Moineddin R, Urbach DR, Rabeneck L and Bierman AS. Trends and inequities in colorectal cancer screening participation in Ontario, Canada, 2005-2011. Cancer Epidemiol 2013;37:946–956. DOI: 10.1016/j.canep.2013.04.007 (cit. on p. 33).

153. Banks E, Beral V, Cameron R et al. Comparison of various characteristics of women who do and do not attend for breast cancer screening. Breast Cancer Res 2001;4:R1. DOI: 10.1186/bcr418. (Cit. on p. 33).

154. Moffat J, Bentley A, Ironmonger L, Boughey A, Radford G and Duffy S. The impact of national cancer awareness campaigns for bowel and lung cancer symptoms on sociodemographic inequalities in immediate key symptom awareness and GP attendances. Br J Cancer s1 2015;112:S14–S21. DOI: 10.1038/bjc.2015.31. (Cit. on p. 33).

155. Weinstein MC, O'Brien B, Hornberger J et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ispor task force on good research practices—modeling studies. Value Health 2003;6:9–17. DOI: 10.1046/j.1524-4733.2003.00234.x. (Cit. on p. 34).

156. Epstein JM. Why model? J Artif Soc Soc Simul 2008;11:12. Available at: http://jasss.soc.surrey.ac.uk/11/4/12.html (cit. on pp. 34, 63, 97, 168, 169).

157. Rasmussen-Torvik LJ, Shay CM, Abramson JG et al. Ideal cardiovascular health is inversely associated with incident cancer the atherosclerosis risk in communities study. Circulation 2013;127:1270–1275. DOI: 10.1161/CIRCULATIONAHA.112.001183. (Cit. on pp. 34, 167).

158. Petticrew M and Roberts H. Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health 2003;57:527–529. DOI: 10.1136/jech.57.7.527. (Cit. on p. 34).

159. Mebius A. Corroborating evidence-based medicine. J Eval Clin Pract 2014;20:915–920. Available at: https://liverpool.idm.oclc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000350450200027&site=eds-live&scope=site (accessed on 30th July 2016) (cit. on p. 35).

160. Howick J, Glasziou P and Aronson J. Evidence-based mechanistic reasoning. J R Soc Med 2010;103:433–441. DOI: 10.1258/jrsm.2010.100146. (Cit. on p. 35).

161. Webber L, Mytton OT, Briggs ADM et al. The Brighton declaration: the value of non-communicable disease modelling in population health sciences. Eur J Epidemiol 2014;29:1–4. DOI: 10.1007/s10654-014-9978-0. (Cit. on pp. 35, 39, 66, 164).

162. Smith BT, Smith PM, Harper S, Manuel DG and Mustard CA. Reducing social inequalities in health: the role of simulation modelling in chronic disease epidemiology to evaluate the impact of population health interventions. J Epidemiol Community Health 2013;68:384–389. DOI: 10.1136/jech-2013-202756. (Cit. on p. 35).

163. Galea S, Riddle M and Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol 2010;39:97–106. DOI: 10.1093/ije/dyp296. (Cit. on p. 35).

164. Ness RB, Koopman JS and Roberts MS. Causal system modeling in chronic disease epidemiology: a proposal. Ann Epidemiol 2007;17:564–568. DOI: 10.1016/j.annepidem.2006.10.014. (Cit. on p. 35).

165. Fone D, Hollinghurst S, Temple M et al. Systematic review of the use and value of computer simulation modelling in population health and health care delivery. J Public Health 2003;25:325–335. DOI: 10.1093/pubmed/fdg075. (Cit. on pp. 35, 36).

166. Unal B, Capewell S and Critchley JA. Coronary heart disease policy models: a systematic review. BMC Public Health 2006;6:213. DOI: 10.1186/1471-2458-6-213. (Cit. on p. 35).

167. Capewell S, Allender S, Critchley J et al. Modelling the UK cardiovascular burden to 2020. British Heart Foundation and Cardiovascular Coalition, 2008. Available at: http://publications.bhf.org.uk/BHFOnline/publication_view.aspx?ps=1000626, (accessed on 16th Oct. 2016) (cit. on p. 35).

168. Speybroeck N, Van Malderen C, Harper S, Müller B and Devleesschauwer B. Simulation models for socioeconomic inequalities in health: a systematic review. Int J Environ Res Public Health 2013;10:5750–5780. DOI: 10.3390/ijerph10115750. (Cit. on p. 35).

169. Murray CJ and Lopez AD. On the Comparable Quantification of Health Risks: Lessons from the Global Burden of Disease Study. Epidemiology (Cambridge, Mass.) 1999;10:594–605 (cit. on p. 36).

170. Kontis V, Mathers CD, Rehm J et al. Contribution of six risk factors to achieving the 25×25 non-communicable disease mortality reduction target: a modelling study. Lancet 2014;384:427–437. DOI: 10.1016/S0140-6736(14)60616-4. (Cit. on pp. 36, 55).

171. Parkin DM, Boyd L and Walker LC. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br J Cancer S2 2011;105:S77–S81. DOI: 10.1038/bjc.2011.489. (Cit. on pp. 36, 55, 67).

172. Scarborough P, Nnoaham KE, Clarke D, Capewell S and Rayner M. Modelling the impact of a healthy diet on cardiovascular disease and cancer mortality. J Epidemiol Community Health 2012;66:420–426. DOI: 10.1136/jech.2010.114520. (Cit. on pp. 37, 164, 174).

173. Levy DT, Huang AT, Currie LM and Clancy L. The benefits from complying with the framework convention on tobacco control: a SimSmoke analysis of 15 European nations. Health Policy Plan 2014;29:1031–1042. DOI: 10.1093/heapol/czt085 (cit. on pp. 37, 124, 148, 160, 195).

174. Briggs ADM, Mytton OT, Kehlbacher A et al. Health impact assessment of the UK soft drinks industry levy: a comparative risk assessment modelling study. The Lancet Public Health 2017;2:e15–e22. DOI: 10.1016/S2468-2667(16)30037-8. (Cit. on p. 37).

175. Scholes S, Bajekal M, Norman P et al. Quantifying policy options for reducing future coronary heart disease mortality in England: a modelling study. PLOS One 2013;8. Ed. by Hernandez A:e69935. DOI: 10.1371/journal.pone.0069935. (Cit. on pp. 37, 38, 131).

176. Gillespie DOS, Allen K, Guzman-Castillo M et al. The health equity and effectiveness of policy options to reduce dietary salt intake in England: policy forecast. PLOS One 2015;10. Ed. by Divaris K:e0127927. DOI: 10.1371/journal.pone.0127927. (Cit. on pp. 37, 38, 123, 135, 144, 195).

177. Levy DT, Cummings KM and Hyland A. Increasing taxes as a strategy to reduce cigarette use and deaths: results of a simulation model. Prev Med 2000;31:279–286. DOI: 10.1006/pmed.2000.0696. (Cit. on p. 37).

178. Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB and Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: the Coronary Heart Disease Policy Model. Am J Public Health 1987;77:1417–1426. DOI: 10.2105/AJPH.77.11.1417. (Cit. on p. 37).

179. Lhachimi SK, Nusselder WJ, Smit HA et al. DYNAMO-HIA – A dynamic modeling tool for generic health impact assessments. PLOS One 2012;7. Ed. by Pai NP:e33317. DOI: 10.1371/journal.pone.0033317. (Cit. on p. 37).

180. Unal B, Critchley JA and Capewell S. Explaining the decline in coronary heart disease mortality in England and Wales between 1981 and 2000. Circulation 2004;109:1101–1107. DOI: 10.1161/01.CIR.0000118498.35499.B2. (Cit. on p. 38).

181. Schlessinger L and Eddy DM. Archimedes: a new model for simulating health care systems—the mathematical formulation. J Biomed Inform 2002;35:37–50. DOI: 10.1016/S1532-0464(02)00006-0. (Cit. on p. 38).

182. Wolfson MC. POHEM–a framework for understanding and modelling the health of human populations. World Health Stat Q 1994;47:157–176 (cit. on p. 38).

183. Hennessy DA, Flanagan WM, Tanuseputro P et al. The Population Health Model (POHEM): an overview of rationale, methods and applications. Popul Health Metr 2015;13:24. DOI: 10.1186/s12963-015-0057-x. (Cit. on p. 38).

184. National Cancer Institute. Cancer Intervention and Surveillance Modeling Network (CISNET). 2016. Available at: http://cisnet.cancer.gov/ (accessed on 2nd Aug. 2016) (cit. on p. 38).

185. National Cancer Institute. CISNET model registry. 2016. Available at: https://resources.cisnet.cancer.gov/registry/packages/ (accessed on 26th Apr. 2016) (cit. on p. 38).

186. Caro JJ, Briggs AH, Siebert U and Kuntz KM. Modeling good research practices—overview: a report of the ISPOR-SMDM modeling good research practices task force-1. Value Health 2012;15:796–803. DOI: 10.1016/j.jval.2012.06.012. (Cit. on pp. 41, 164).

187. Siebert U, Alagoz O, Bayoumi AM et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. Value Health 2012;15:812–820. DOI: 10.1016/j.jval.2012.06.014. (Cit. on p. 41).

188. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J and Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM modeling good research practices task force-4. Value Health 2012;15:821–827. DOI: 10.1016/j.jval.2012.04.013. (Cit. on p. 41).

189. Pitman R, Fisman D, Zaric GS et al. Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-5. Value Health 2012;15:828–834. DOI: 10.1016/j.jval.2012.06.011. (Cit. on p. 42).

190. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P and Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force-2. Value Health 2012;15:804–811. DOI: 10.1016/j.jval.2012.06.016. (Cit. on p. 42).

191. Williamson P. The role of the International Journal of Microsimulation. International Journal of Microsimulation 2007;1:1–2. Available at: http://microsimulation.org/IJM/V1_1/IJM_1_1_1.pdf (accessed on 21st July 2014) (cit. on p. 42).

192. Orcutt GH. A new type of socio-economic system. Rev Econ Stat 1957;39:116. DOI: 10.2307/1928528. (Cit. on p. 42).

193. Troitzsch KG. Social science microsimulation. Ed. by Troitzsch KG, Mueller U, Gilbert GN and Doran JE. Springer Science & Business Media, 1996. 492 pp. DOI: 10.1007/978-3-662-03261-9 (cit. on p. 42).

194. Glied S and Tilipman N. Simulation modeling of health care policy. Annu Rev Public Health 2010;31:439–455. DOI: 10.1146/annurev.publhealth.012809.103542. (Cit. on p. 42).

195. Rutter CM, Miglioretti DL and Savarino JE. Evaluating risk factor assumptions: a simulation-based approach. BMC Med Inform Decis Mak 2011;11:55. DOI: 10.1186/1472-6947-11-55. (Cit. on p. 42).

196. Zucchelli E, Jones AM and Rice N. The evaluation of health policies through dynamic microsimulation methods. International Journal of Microsimulation 2012;5:2–20. Available at: http://ima.natsem.canberra.edu.au/IJM/V5_1/1_IJM_5_1_spring_2012_Zuchelli_Rice_Jones.pdf (accessed on 21st July 2014) (cit. on pp. 42, 43, 61).

197. McLennan D, Barnes H, Noble M, Davies J, Garratt E and Dibben C. The English indices of deprivation 2010. Tech. rep. Department for Communities and Local Government, 2011. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6871/1871208.pdf (accessed on 4th Aug. 2016) (cit. on p. 43).

198. Mindell J, Biddulph JP, Hirani V et al. Cohort profile: the Health Survey for England. Int J Epidemiol 2012;41:1585–1593. DOI: 10.1093/ije/dyr199. (Cit. on p. 46).

199. Müller K and Axhausen KW. Population synthesis for microsimulation: State of the art. ETH Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau (IVT), 2010. Available at: https://www.researchgate.net/profile/Kay_Axhausen/publication/228973867_Population_synthesis_for_microsimulation_State_of_the_art/links/0deec517bbebb9606c000000.pdf (accessed on 11th June 2016) (cit. on pp. 46, 47).

200. Camitz M, Brouwers L, Tegnell A and Boman M. Economic consequences to society of pandemic h1N1 influenza 2009: preliminary results for Sweden. Euro Surveill 2009;14 (cit. on p. 46).

201. Alfons A, Kraft S, Templ M and Filzmoser P. Simulation of close-to-reality population data for household surveys with application to EU-SILC. Stat Methods Appt 2011;20:383–407. DOI: 10.1007/s10260-011-0163-2. (Cit. on p. 47).

202. Alfons A, Templ M and Filzmoser P. Simulation of EU-SILC population data: using the R package simPopulation. Research Report CS-2010-5, Department of Statistics and Probability Theory, Vienna University of Technology, 2010. Available at: http://www.statistik.tuwien.ac.at/forschung/CS/CS-2010-5complete.pdf (accessed on 30th Oct. 2013) (cit. on p. 47).

203. Frazier TJ and Alfons A. Generating a close-to-reality synthetic population of Ghana. SSRN Scholarly Paper ID 2086345. Rochester, NY: Social Science Research Network, 2012. DOI: 10.2139/ssrn.2086345. (Cit. on p. 47).

204. NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2011 [computer file]. Colchester, Essex: UK Data Archive [distributor]. 2013. (Cit. on pp. 48, 50, 70, 104).

205.   NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2012 [computer file]. Colchester, Essex: UK Data Archive [distributor]. 2014. (Cit. on pp. 48, 50, 70, 104).

206.   NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2005 [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor]. 2011. (Cit. on pp. 48, 50, 70).

207.   NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2006 [computer file]. 4th Edition. Colchester, Essex: UK Data Archive [distributor]. 2011. (Cit. on pp. 48, 50, 58, 70).

208.   NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2007 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. 2010. (Cit. on pp. 48, 50, 70).

209.   R Core Team. R: a language and environment for statistical computing. Version 3.3.0. Vienna, Austria: R Foundation for Statistical Computing, 2015. Available at: https://www.R-project.org/ (cit. on pp. 48, 66, 97, 106).

210.   Alfons A and Kraft S. Simpopulation: simulation of synthetic populations for surveys based on sample data. Version 0.4.1. 2013. Available at: http://CRAN.R-project.org/package=simPopulation (cit. on p. 48).

211.   Meindl B, Templ M, Alfons A, Kowarik A and Mathieu Ribatet with contributions from. Simpop: simulation of synthetic populations for survey data considering auxiliary information. Version 0.3.0. 2016. Available at: http://CRAN.R-project.org/package=simPop (cit. on p. 48).

212.   Giles Horsfield SSD. Chapter 3: equivalised Income. Office for National Statistics. 2012. Available at: http://www.ons.gov.uk/ons/rel/family-spending/family-spending/family-spending-2012-edition/art-chapter-3--equivalised-income.html (accessed on 4th May 2015) (cit. on p. 48).

213.   Economic Labour and Social Analysis Office for National Statistics. SOC2010 volume 3: the National Statistics socio-economic classification (NS-SEC rebased on the SOC2010). Office for National Statistics. NS-SEC has been constructed to measure the employment relations and conditions of occupations. 2010. Available at: http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html (accessed on 6th Jan. 2015) (cit. on p. 48).

214.   Ji C, Sykes L, Paul C et al. Systematic review of studies comparing 24-hour and spot urine collections for estimating population salt intake. Rev Panam Salud Publica 2012;32:307–315. DOI: 10.1590/S1020-49892012001000010. (Cit. on pp. 49, 51).

215.   Ji C, Miller MA, Venezia A, Strazzullo P and Cappuccio FP. Comparisons of spot vs 24-h urine samples for estimating population salt intake: Validation study in two independent samples of adults in Britain and Italy. Nutr Metab Cardiovasc Dis 2014;24:140–147. DOI: 10.1016/j.numecd.2013.06.011. (Cit. on pp. 49, 51, 145).

216. Mulrow CD, Chiquette E, Angel L et al. Dieting to reduce body weight for controlling hypertension in adults. Cochrane Database Syst Rev 4 2000:CD000484. DOI: 10.1002/14651858.cd000484.pub2 (cit. on p. 49).

217. Goldstein DJ. Beneficial health effects of modest weight loss. Int J Obes Relat Metab Disord 1992;16:397–415 (cit. on p. 49).

218. MacMahon S, Cutler J, Brittain E and Higgins M. Obesity and hypertension: epidemiological and clinical issues. Eur Heart J 1987;8:57–70. DOI: 10.1093/eurheartj/8.suppl_b.57 (cit. on p. 49).

219. Dyer AR and Elliott P. The INTERSALT study: relations of body mass index to blood pressure. INTERSALT Co-operative Research Group. J Hum Hypertens 1989;3:299–308 (cit. on pp. 49, 51, 183).

220. Dattilo AM and Kris-Etherton PM. Effects of weight reduction on blood lipids and lipoproteins: a meta-analysis. Am J Clin Nutr 1992;56:320–328 (cit. on p. 49).

221. Kahn SE, Hull RL and Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature 2006;444:840–846. DOI: 10.1038/nature05482 (cit. on p. 49).

222. Lahti-Koski M, Pietinen P, Heliövaara M and Vartiainen E. Associations of body mass index and obesity with physical activity, food choices, alcohol intake, and smoking in the 1982–1997 FINRISK Studies. Am J Clin Nutr 2002;75:809–817. Available at: http://ajcn.nutrition.org/content/75/5/809 (accessed on 16th Oct. 2015) (cit. on p. 49).

223. DiPietro L. Physical activity in the prevention of obesity: current evidence and research issues. Med Sci Sports Exerc 1999;31:S542–546. DOI: 10.1097/00005768-199911001-00009 (cit. on p. 49).

224. Fogelholm M and Kukkonen-Harjula K. Does physical activity prevent weight gain–a systematic review. Obes Rev 2000;1:95–111. DOI: 10.1046/j.1467-789x.2000.00016.x (cit. on p. 49).

225. Dochi M, Sakata K, Oishi M, Tanaka K, Kobayashi E and Suwazono Y. Smoking as an independent risk factor for hypertension: a 14-year longitudinal study in male Japanese workers. Tohoku J Exp Med 2009;217:37–43. Available at: http://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=mnh&AN=19155606&site=ehost-live&scope=site (accessed on 16th Oct. 2015) (cit. on p. 49).

226. Bowman TS, Gaziano JM, Buring JE and Sesso HD. A prospective study of cigarette smoking and risk of incident hypertension in women. J Am Coll Cardiol 2007;50:2085–2092. DOI: 10.1016/j.jacc.2007.08.017. (Cit. on p. 49).

227. Cleveland WS, Devlin SJ and Grosse E. Regression by local fitting: methods, properties, and computational algorithms. J Econom 1988;37:87–114 (cit. on p. 50).

228. Deville JC and Särndal CE. Calibration estimators in survey sampling. J Am Stat Assoc 1992;87:376–382. DOI: 10.2307/2290268 (cit. on p. 50).

229. Deville JC, Särndal CE and Sautory O. Generalized raking procedures in survey sampling. J Am Stat Assoc 1993;88:1013–1020. DOI: 10.2307/2290793 (cit. on p. 50).

230. Lumley T. Survey: analysis of complex survey samples. Version 3.29-5. 2013. Available at: http://cran.r-project.org/web/packages/survey/index.html (cit. on pp. 50, 106).

231. Lumley TS. Complex surveys: a guide to analysis using R. John Wiley & Sons, 2011. 296 pp. Available at: http://www.ebook.de/de/product/21162302/thomas_s_lumley_complex_surveys.html (cit. on p. 50).

232. NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2001 [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor]. 2010. (Cit. on pp. 50, 70).

233. NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2002 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. 2010. (Cit. on pp. 50, 70).

234. NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2003 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. 2010. (Cit. on pp. 50, 70).

235. NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2004 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. 2010. (Cit. on pp. 50, 70).

236. NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2008 [computer file]. 4th Edition. Colchester, Essex: UK Data Archive [distributor]. 2013. (Cit. on pp. 50, 70).

237. NatCen Social Research and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2009 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. 2011. (Cit. on pp. 50, 70).

238. NatCen Social Research, Royal Free and University College London. Department of Epidemiology and Public Health. Health Survey for England, 2010 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. 2012. (Cit. on pp. 50, 70).

239. Brown IJ, Dyer AR, Chan Q et al. Estimating 24-hour urinary sodium excretion from casual urinary sodium concentrations in Western populations the INTERSALT study. Am J Epidemiol 2013;177:1180–1192. DOI: 10.1093/aje/kwt066. (Cit. on pp. 51, 183).

240. An assessment of dietary sodium levels among adults (aged 19-64) in the UK general population in 2008, based on analysis of dietary sodium in 24-hour urine samples. Research rep. National Centre for Social Research, 2008. Available at: http://tna.europarchive.org/20110116113217/http://www.food.gov.uk/multimedia/pdfs/08sodiumreport.pdf (cit. on pp. 51, 137, 226, 227).

241. An assessment of dietary sodium levels among adults (aged 19-64) in the general population in England, based on analysis of dietary sodium in 24-hour urine samples. National Centre for Social Research, 2006. Available at: http://tna.europarchive.org/20110116113217/http://www.food.gov.uk/multimedia/pdfs/englandsodiumreport.pdf (cit. on pp. 51, 137, 226, 227).

242. Henderson L, Irving K, Gregory J et al. National Diet and Nutrition Survey: adults aged 19 to 64 years. Volume 3: vitamin and mineral intake and urinary analytes. Research rep. London: The Stationery Office, 2003. Available at: http://www.ons.gov.uk/ons/rel/lifestyles/the-national-diet-and-nutrition-survey/2001-edition/the-national-diet-and-nutrition-survey---volume-3.pdf (accessed on 5th June 2015) (cit. on pp. 51, 137, 184, 226, 227).

243. Hippisley-Cox J, Coupland C and Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ Open 2014;4:e005809. DOI: 10.1136/bmjopen-2014-005809. (Cit. on p. 54).

244. Hippisley-Cox J, Coupland C, Robson J, Sheikh A and Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ 2009;338:b880. DOI: 10.1136/bmj.b880. (Cit. on p. 54).

245. Office for National Statistics. National population projections, 2012. 2013. Available at: http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2012-based-projections/index.html (accessed on 11th Nov. 2014) (cit. on pp. 56, 70).

246. Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum 1953;9:531–541 (cit. on p. 57).

247. Barendregt JJ, van Oortmarssen GJ, Vos T and Murray CJL. A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. Popul Health Metr 2003;1:4. DOI: 10.1186/1478-7954-1-4. (Cit. on p. 58).

248. Lim SS, Vos T, Flaxman AD et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2224–2260. DOI: 10.1016/S0140-6736(12)61766-8. (Cit. on p. 58).

249. Heidrich J, Wellmann J, Heuschmann PU, Kraywinkel K and Keil U. Mortality and morbidity from coronary heart disease attributable to passive smoking. Eur Heart J 2007;28:2498–2502. DOI: 10.1093/eurheartj/ehm151. (Cit. on p. 58).

250. McLean S, Barbour V, Wild S, Simpson C and Sheikh A. Models for estimating projections for disease prevalence and burden: a systematic review focusing on chronic obstructive pulmonary disease. J Health Serv Res Policy 2015;20:246–253. DOI: 10.1177/1355819615579232. (Cit. on p. 58).

251. Pichon-Riviere A, Augustovski F, Bardach A, Colantonio L and Latinclen Tobacco
      Research Group. Development and validation of a microsimulation economic model
      to evaluate the disease burden associated with smoking and the cost-effectiveness
      of tobacco control interventions in Latin America. Value Health 2011;14:51–9. DOI:
      10.1016/j.jval.2011.05.010. (Cit. on p. 58).

252. Boshuizen HC, Lhachimi SK, Baal PHM van et al. The DYNAMO-HIA Model: an
      efficient implementation of a risk factor/chronic disease Markov model for use in
      health impact assessment (HIA). Demography 2012;49:1259–1283. DOI: 10.1007/s13
      524-012-0122-z. (Cit. on p. 58).

253. Blakely T, Cobiac LJ, Cleghorn CL et al. Health, health inequality, and cost impacts of
      annual increases in tobacco tax: multistate life table modeling in New Zealand. PLOS
      Med 2015;12. Ed. by Novotny TE:e1001856. DOI: 10.1371/journal.pmed.1001856. (Cit.
      on pp. 58, 165).

254. Webber L, Divajeva D, Marsh T et al. The future burden of obesity-related diseases
      in the 53 WHO European-Region countries and the impact of effective interventions:
      a modelling study. BMJ Open 2014;4:e004787. DOI: 10.1136/bmjopen-2014-004787.
      (Cit. on p. 58).

255. Saidi O, O'Flaherty M, Mansour NB et al. Forecasting Tunisian type 2 diabetes pre-
      valence to 2027: validation of a simple model. BMC Public Health 2015;15:104. DOI:
      10.1186/s12889-015-1416-z (cit. on p. 58).

256. Office for National Statistics. Published ad hoc data: health, requests during Decem-
      ber 2014. 2014. Available at: http://www.ons.gov.uk/ons/about-ons/business-
      transparency/freedom-of-information/what-can-i-request/published-ad-hoc-
      data/health/december-2014/number-of-registered-deaths-by-sex--cause--year--
      the-adjusted-index.xls (accessed on 14th Oct. 2015) (cit. on pp. 58–60, 70).

257. Office for National Statistics. Cancer registrations in England. Cancer Registrations
      in England. 2015. Available at: http://www.ons.gov.uk/ons/rel/vsob1/cancer-
      registrations-in-england/index.html (accessed on 28th Sept. 2015) (cit. on p. 59).

258. Office for National Statistics. Cancer survival rates - cancer survival in England: pa-
      tients diagnosed 2006–2010 and followed up to 2011. Cancer Survival Rates - Cancer
      Survival in England: Patients Diagnosed. 2012. Available at: http://www.ons.gov.uk/
      ons/rel/cancer-unit/cancer-survival/2006---2010--followed-up-to-2011/index.html
      (accessed on 28th Sept. 2015) (cit. on p. 59).

259. Hyndman RJ and Shahid Ullah M. Robust forecasting of mortality and fertility rates:
      a functional data approach. Comput Stat Data Anal 2007;51:4942–4956. DOI: 10.1016/
      j.csda.2006.07.028. (Cit. on p. 60).

260. Hyndman RJ. Demography: forecasting mortality, fertility, migration and popula-
      tion data. In collab. with Booth H, Tickle L and Maindonald J. Version 1.18. 2014.
      Available at: http://CRAN.R-project.org/package=demography (cit. on p. 60).

261.   Siemerink EJM, Hospers GAP, Mulder NH, Siesling S and van der Aa MA. Disparities in survival of stomach cancer among different socioeconomic groups in North-East Netherlands. Cancer Epidemiol 2011;35:413–416. DOI: 10.1016/j.canep.2011.02.015. (Cit. on pp. 60, 137).

262.   The DECODE Study Group. Is the current definition for diabetes relevant to mortality risk from all causes and cardiovascular and noncardiovascular diseases? Diabetes Care 2003;26:688–696. DOI: 10.2337/diacare.26.3.688. (Cit. on pp. 60, 67, 74).

263.   Doll R, Peto R, Boreham J and Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ 2004;328:1519. DOI: 10.1136/bmj. 38142.554479.AE. (Cit. on pp. 60, 67, 71).

264.   Koerkamp BG, Stijnen T, Weinstein MC and Hunink MGM. The combined analysis of uncertainty and patient heterogeneity in medical decision models. Med Decis Making 2011;31:650–661. DOI: 10.1177/0272989X10381282. (Cit. on pp. 62, 63).

265.   Koerkamp BG, Weinstein MC, Stijnen T, Heijenbrok-Kal MH and Hunink MGM. Uncertainty and patient heterogeneity in medical decision models. Med Decis Making 2010;30:194–205. DOI: 10.1177/0272989X09342277. (Cit. on pp. 62, 63).

266.   Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ and Paltiel AD. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM modeling good research practices Task Force-6. Value Health 2012;15:835–842. DOI: 10. 1016/j.jval.2012.04.014. (Cit. on pp. 62, 63).

267.   Clark CE. Letter to the editor-the PERT model for the distribution of an activity time. Oper Res 1962;10:405–406. DOI: 10.1287/opre.10.3.405 (cit. on p. 63).

268.   Briggs AH, Claxton K and Sculpher MJ. Decision modelling for health economic evaluation. Oxford University Press, 2006. 260 pp. (cit. on p. 63).

269.   Bross IDJ. How to use ridit analysis. Biometrics 1958;14:18–38. DOI: 10.2307/2527727 (cit. on p. 65).

270.   Dowle M, Short T, Lianoglou S and Srinivasan A. Data.table: extension of data.frame. In collab. with Saporta R and Antonyan E. Version 1.9.6. 2015. Available at: https: //github.com/Rdatatable/data.table/ (cit. on p. 66).

271.   Revolution Analytics and Weston S. Foreach: foreach looping construct for R. Version 1.4.2. 2014. Available at: http://CRAN.R-project.org/package=foreach (cit. on p. 66).

272.   Gaujoux R. doRNG: generic reproducible parallel backend for foreach loops. Version 1.6. 2014. Available at: http://CRAN.R-project.org/package=doRNG (cit. on p. 66).

273.   L'Ecuyer P. Good parameters and implementations for combined multiple recursive random number generators. Oper Res 1999;47:159–164. DOI: 10.1287/opre.47.1.159. (Cit. on p. 66).

274. Rockhill B, Newman B and Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998;88:15–19. DOI: 10.2105/AJPH.88.1.15. (Cit. on p. 67).

275. Bruzzi P, Green SB, Byar DP, Brinton LA and Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol 1985;122:904–914. Available at: http://aje.oxfordjournals.org/content/122/5/904 (accessed on 9th July 2014) (cit. on p. 67).

276. Laaksonen MA, Virtala E, Knekt P, Oja H and Härkänen T. SAS macros for calculation of population attributable fraction in a cohort study design. J Stat Softw i07 2011;43. DOI: 10.18637/jss.v043.i07. (Cit. on p. 67).

277. Rückinger S, Kries R von and Toschke AM. An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. BMC Med Res Methodol 2009;9:7. DOI: 10.1186/1471-2288-9-7. (Cit. on pp. 67, 68).

278. González CA, Pera G, Agudo A et al. Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). Int J Cancer 2003;107:629–634. DOI: 10.1002/ijc.11426. (Cit. on pp. 71, 72, 193).

279. Huxley RR and Woodward M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. Lancet 2011;378:1297–1305. DOI: 10.1016/S0140-6736(11) 60781-2. (Cit. on pp. 71, 187).

280. He J, Vupputuri S, Allen K, Prerost MR, Hughes J and Whelton PK. Passive smoking and the risk of coronary heart disease — A meta-analysis of epidemiologic studies. N Engl J Med 1999;340:920–926. DOI: 10.1056/NEJM199903253401204. (Cit. on pp. 72, 188).

281. Oono IP, Mackay DF and Pell JP. Meta-analysis of the association between second-hand smoke exposure and stroke. J Public Health 2011;33:496–502. DOI: 10.1093/pubmed/fdr025. (Cit. on pp. 72, 191).

282. Kim CH, Lee YCA, Hung RJ et al. Exposure to secondhand tobacco smoke and lung cancer by histological type: a pooled analysis of the International Lung Cancer Consortium (ILCCO). Int J Cancer 2014;135:1918–1930. DOI: 10.1002/ijc.28835. (Cit. on pp. 72, 193).

283. Taylor R, Najafi F and Dobson A. Meta-analysis of studies of passive smoking and lung cancer: effects of study type and continent. Int J Epidemiol 2007;36:1048–1059. DOI: 10.1093/ije/dym158. (Cit. on p. 72).

284. The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010;375:2215–2222. DOI: 10.1016/S0140-6736(10)60484-9. (Cit. on pp. 73, 189, 192).

285.   Dauchet L, Amouyel P, Hercberg S and Dallongeville J. Fruit and vegetable consumption and risk of coronary heart disease: a meta-analysis of cohort studies. J Nutr 2006;136:2588–2593. Available at: http://jn.nutrition.org/content/136/10/2588 (accessed on 13th May 2014) (cit. on pp. 74, 190).

286.   Dauchet L, Amouyel P and Dallongeville J. Fruit and vegetable consumption and risk of stroke: a meta-analysis of cohort studies. Neurology 2005;65:1193–1197. DOI: 10.1212/01.wnl.0000180600.09719.53. (Cit. on pp. 74, 193).

287.   Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM and Wong JB. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force-7. Value Health 2012;15:843–850. DOI: 10.1016/j.jval.2012.04.012. (Cit. on pp. 77, 164).

288.   Vemer P, Krabbe PFM, Feenstra TL, Van Voorn GAK, Ramos C and Al MJ. Improving model validation in health technology assessment: comments on guidelines of the ISPOR-SMDM modeling good research practices task force. Value Health 2013;16:1106–1107. DOI: 10.1016/j.jval.2013.06.015. (Cit. on p. 77).

289.   Friendly M. Mosaic displays for multi-way contingency tables. J Am Stat Assoc 1994;89:190–200. DOI: 10.2307/2291215 (cit. on p. 78).

290.   Cox NR. Estimation of the correlation between a continuous and a discrete variable. Biometrics 1974;30:171–178. Available at: http://www.jstor.org/stable/2529626 (cit. on p. 78).

291.   Olsson U. Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika 1979;44:443–460. DOI: 10.1007/bf02296207 (cit. on p. 78).

292.   Bailey AEP̃. wCorr: weighted correlations. Version 1.8.0. 2016. Available at: https://CRAN.R-project.org/package=wCorr (cit. on p. 78).

293.   Schmid VJ and Held L. Bayesian age-period-cohort modeling and prediction-BAMP. J Stat Software 2007;21:1–15. Available at: http://www.doaj.org/doaj?func=fulltext&aId=295518 (accessed on 3rd Oct. 2013) (cit. on pp. 78, 86, 87, 89, 90, 92, 93, 95, 96, 98).

294.   Pearson-Stuttard J, Guzman Castillo M, Penalvo JL et al. Modelling future cardiovascular disease mortality in the United States: national trends and racial and ethnic disparities. Circulation 2016;133:967–78. DOI: 10.1161/CIRCULATIONAHA.115.019904 (cit. on p. 78).

295.   Tammemägi MC, Church TR, Hocking WG et al. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. PLOS Med 2014;11. Ed. by Massad M:e1001764. DOI: 10.1371/journal.pmed.1001764. (Cit. on p. 79).

296.   De Matteis S, Consonni D, Pesatori AC et al. Are women who smoke at higher risk for lung cancer than men who smoke? Am J Epidemiol 2013;177:601–612. DOI: 10.1093/aje/kws445. (Cit. on p. 79).

297.   Office for National Statistics. Results of the ICD-10 v2010 bridge coding study, England and Wales. 2011. Available at: http://www.ons.gov.uk/ons/rel/subnational-health3/results-of-the-icd-10-v2010-bridge-coding-study--england-and-wales--2009/2009/index.html (accessed on 21st Sept. 2015) (cit. on pp. 84, 88).

298.   Kypridemos C, Allen K, Hickey GL et al. Cardiovascular screening to reduce the burden from cardiovascular disease: microsimulation study to quantify policy options. BMJ 2016;353:i2793. DOI: 10.1136/bmj.i2793. (Cit. on pp. 97, 172).

299.   Kypridemos C, Bandosz P, Hickey G et al. OP04 Quantifying the contribution of statins to the decline in population mean cholesterol by socioeconomic group in England 1991–2012: a modelling study. J Epidemiol Community Health Suppl 1 2015;69:A10–A10. DOI: 10.1136/jech-2015-206256.4. (Cit. on pp. 97, 172).

300.   Kypridemos C, Allen K, Bandosz P et al. OP10 Quantifying national policy options for equitable primary cardiovascular disease prevention in England up to 2025: a micro-simulation study. J Epidemiol Community Health Suppl 1 2015;69:A13. DOI: 10.1136/jech-2015-206256.10. (Cit. on pp. 97, 172).

301.   Kypridemos C, Guzman-Castillo M, Hyseni L et al. An opportunity to reduce the burdens and inequalities in cardiovascular disease and gastric cancer caused by dietary salt: IMPACT$_{NCD}$ microsimulation study. Lancet 2015. Public Health Science November 2015 386, Supplement 2:S47. DOI: 10.1016/S0140-6736(15)00885-5. (Cit. on pp. 97, 172).

302.   Grajales CAG. Created by statisticians for statisticians: how R took the world of statistics by storm - Statistics Views. 2015. Available at: http://www.statisticsviews.com/details/feature/8585391/Created-by-statisticians-for-statisticians-How-R-took-the-world-of-statistics-by.html (accessed on 19th Aug. 2016) (cit. on p. 97).

303.   Morton GM, Lee SM, Buss DH and Lawrance P. Intakes and major dietary sources of cholesterol and phytosterols in the British diet. J Hum Nutr Diet 1995;8:429–440. DOI: 10.1111/j.1365-277X.1995.tb00338.x. (Cit. on p. 103).

304.   Department of Health. Pharmacies and prescriptions: statistics. 2014. Available at: http://webarchive.nationalarchives.gov.uk/20120503222906/http:/www.dh.gov.uk/en/PublicationsAndStatistics/Statistics/StatisticalWorkAreas/StatisticalHealthCare/DH_4086488 (accessed on 15th June 2014) (cit. on p. 103).

305.   Farzadfar F, Finucane MM, Danaei G et al. National, regional, and global trends in serum total cholesterol since 1980: systematic analysis of health examination surveys and epidemiological studies with 321 country-years and 3·0 million participants. Lancet 2011;377:578–586. DOI: 10.1016/S0140-6736(10)62038-7. (Cit. on pp. 103, 132).

306.   Cohen JD, Cziraky MJ, Cai Q et al. 30-year trends in serum lipids among United States adults: results from the National Health and Nutrition Examination Surveys II, III, and 1999–2006. Am J Cardiol 2010;106:969–975. DOI: 10.1016/j.amjcard.2010.05.030. (Cit. on p. 103).

307.   Johnson CL, Rifkind BM, Sempos CT et al. Declining serum total cholesterol levels among US adults: the National Health and Nutrition Examination Surveys. JAMA 1993;269:3002–3008. DOI: 10.1001/jama.1993.03500230084034. (Cit. on p. 103).

308.   Capewell S and Ford ES. Why have total cholesterol levels declined in most developed countries? BMC Public Health 2011;11:641. DOI: 10.1186/1471-2458-11-641. (Cit. on pp. 103, 105, 117).

309.   Thorsson B, Steingrimsdottir L, Halldorsdottir S et al. Changes in total cholesterol levels in Western societies are not related to statin, but rather dietary factors: the example of the Icelandic population. Eur Heart J 2013;34:1778–1782. DOI: 10.1093/eurheartj/ehs395. (Cit. on pp. 103, 116, 171).

310.   Eliasson M, Janlert U, Jansson JH and Stegmayr B. Time trends in population cholesterol levels 1986–2004: influence of lipid-lowering drugs, obesity, smoking and educational level. The northern Sweden MONICA study. J Intern Med 2006;260:551–559. DOI: 10.1111/j.1365-2796.2006.01730.x. (Cit. on pp. 103, 116, 171).

311.   Cífková R, Škodová Z, Bruthans J et al. Longitudinal trends in major cardiovascular risk factors in the Czech population between 1985 and 2007/8. Czech MONICA and Czech post-MONICA. Atherosclerosis 2010;211:676–681. DOI: 10.1016/j.atherosclerosis.2010.04.007. (Cit. on pp. 103, 116, 171).

312.   Valsta LM, Tapanainen H, Sundvall J et al. Explaining the 25-year decline of serum cholesterol by dietary changes and use of lipid-lowering medication in Finland. Public Health Nutr Special Issue 6A 2010;13:932–938. DOI: 10.1017/S1368980010001126 (cit. on pp. 103, 116, 171).

313.   Scholes S, Bajekal M, Love H et al. Persistent socioeconomic inequalities in cardiovascular risk factors in England over 1994-2008: A time-trend analysis of repeated cross-sectional data. BMC Public Health 2012;12:129. DOI: 10.1186/1471-2458-12-129. (Cit. on pp. 103, 119, 121).

314.   Blackman T. Statins, saving lives, and shibboleths. BMJ 2007;334:902. DOI: 10.1136/bmj.39163.563519.55. (Cit. on p. 103).

315.   Stone NJ, Robinson JG, Lichtenstein AH et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;129:S1–45. DOI: 10.1161/01.cir.0000437738.63853.7a (cit. on pp. 103, 117).

316.   Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. NICE guidelines [CG181]. National Institute for Health and Care Excellence, 2014. Available at: http://www.nice.org.uk/Guidance/CG181 (accessed on 26th Aug. 2014) (cit. on p. 104).

317.   Wise J. Open letter raises concerns about NICE guidance on statins. BMJ 2014;348:3937. DOI: 10.1136/bmj.g3937. (Cit. on p. 104).

318. Thompson R, Gerada C, Haslam D et al. NICE statin letter: concerns about the latest NICE draft guidance on statins. open. 2014. Available at: http://www.nice.org.uk/media/877/%7BAC%7D/NICE_statin_letter.pdf (cit. on pp. 104, 117).

319. Department for Communities and Local Government. English indices of deprivation 2010 - Publications - GOV.UK. 2011. Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010 (accessed on 26th Aug. 2014) (cit. on p. 104).

320. Office of Population Censuses and Surveys. Social Survey Division. Health Survey for England, 1991-1992 : Combined Data File [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. 1997. (Cit. on p. 104).

321. Capewell S, Morrison CE and McMurray JJ. Contribution of modern cardiovascular treatment and risk factor changes to the decline in coronary heart disease mortality in Scotland between 1975 and 1994. Heart 1999;81:380–386. DOI: 10.1136/hrt.81.4.380. (Cit. on p. 105).

322. Korn LE and Graubard IB. Confidence intervals for proportions with small expected number of positive counts estimated from survey data. Surv Methodol 1998;24:193–201. Available at: http://www.statcan.gc.ca/pub/12-001-x/1998002/article/4356-eng.pdf (cit. on p. 106).

323. Law MR, Wald NJ and Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. BMJ 2003;326:1423. DOI: 10.1136/bmj.326.7404.1423. (Cit. on pp. 106, 194).

324. Kong SX, Crawford SY, Gandhi SK et al. Efficacy of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors in the treatment of patients with hypercholesterolemia: a meta-analysis of clinical trials. Clin Ther 1997;19:778–797. DOI: 10.1016/S0149-2918(97)80102-6. (Cit. on pp. 106, 107).

325. Edwards JE and Moore RA. Statins in hypercholesterolaemia: a dose-specific meta-analysis of lipid changes in randomised, double blind trials. BMC Fam Pract 2003;4:18. DOI: 10.1186/1471-2296-4-18. (Cit. on pp. 106, 107, 119, 194).

326. Jones MD P, Kafonek MD S, Laurora PharmD I and Hunninghake MD D. Comparative dose efficacy study of atorvastatin versus simvastatin, pravastatin, lovastatin, and fluvastatin in patients with hypercholesterolemia (the CURVES study). Am J Cardiol 1998;81:582–587. DOI: 10.1016/S0002-9149(97)00965-X. (Cit. on pp. 106, 107).

327. Centre HSCI. Prescription cost analysis 2011. 2012. Available at: http://www.hscic.gov.uk/pubs/prescostanalysis2011 (accessed on 23rd Feb. 2014) (cit. on pp. 106, 113).

328. Health and Social Care Information Centre. Prescription cost analysis 2012. 2013. Available at: http://www.hscic.gov.uk/searchcatalogue?productid=11412&q=prescription+cost+analysis&topics=0%2fPrescribing&sort=Relevance&size=10&page=1#top (accessed on 23rd Feb. 2014) (cit. on pp. 106, 113).

329. Law MR, Morris JK and Wald NJ. Environmental tobacco smoke exposure and ischaemic heart disease: an evaluation of the evidence. BMJ 1997;315:973–980. DOI: 10.1136/bmj.315.7114.973. (Cit. on p. 107).

330. Cochran WG. Sampling techniques. Vol. xiv. Oxford, England: John Wiley, 1953 (cit. on p. 107).

331. Gatz DF and Smith L. The standard error of a weighted mean concentration—I. Bootstrapping vs other methods. Atmos Environ 1995;29:1185–1193. DOI: 10.1016/1352-2310(94)00210-C. (Cit. on p. 107).

332. Ekstrom C. MESS: miscellaneous esoteric statistical script. Version R package version 0.2-1. 2014. Available at: http://CRAN.R-project.org/package=MESS (cit. on p. 108).

333. Armitage P, Berry G and Matthews JNS. Statistical methods in medical research. John Wiley & Sons, 2008 (cit. on p. 108).

334. Bajekal M, Scholes S, Love H et al. Analysing recent socioeconomic trends in coronary heart disease mortality in England, 2000–2007: a population modelling study. PLOS Med 2012;9. Ed. by Vos T:e1001237. DOI: 10.1371/journal.pmed.1001237. (Cit. on p. 116).

335. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M and Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007;335:136. DOI: 10.1136/bmj.39261.471806.55. (Cit. on p. 116).

336. Collins GS and Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;344:e4181. DOI: 10.1136/bmj.e4181. (Cit. on pp. 116, 123, 246).

337. Ashworth M, Lloyd D, Smith RS, Wagner A and Rowlands G. Social deprivation and statin prescribing: a cross-sectional analysis using data from the new UK general practitioner 'Quality and Outcomes Framework'. J Public Health 2007;29:40–47. DOI: 10.1093/pubmed/fdl068. (Cit. on pp. 116, 119).

338. Wu J, Zhu S, Yao GL, Mohammed MA and Marshall T. Patient factors influencing the prescribing of lipid lowering drugs for primary prevention of cardiovascular disease in UK general practice: a national retrospective cohort study. PLOS One 2013;8. Ed. by Baradaran HR:e67611. DOI: 10.1371/journal.pone.0067611. (Cit. on pp. 116, 118).

339. Stocks NP, Ryan P, McElroy H and Allan J. Statin prescribing in Australia: socioeconomic and sex differences. A cross-sectional study. Med J Aust 2004;180:229–231 (cit. on p. 116).

340. Ohlsson H, Lynch K and Merlo J. Is the physician's adherence to prescription guidelines associated with the patient's socio-economic position? An analysis of statin prescription in South Sweden. J Epidemiol Community Health 2010;64:678–683. DOI: 10.1136/jech.2008.081166. (Cit. on p. 116).

341. Thomsen RW, Johnsen SP, Olesen AV et al. Socioeconomic gradient in use of statins among Danish patients: population-based cross-sectional study. Br J Clin Pharmacol 2005;60:534–542. DOI: 10.1111/j.1365-2125.2005.02494.x. (Cit. on p. 116).

342. Franks P, Tancredi D, Winters P and Fiscella K. Cholesterol treatment with statins: Who is left out and who makes it to goal? BMC Health Serv Res 2010;10:68. DOI: 10.1186/1472-6963-10-68. (Cit. on p. 116).

343. Reinikainen J, Laatikainen T, Karvanen J and Tolonen H. Lifetime cumulative risk factors predict cardiovascular disease mortality in a 50-year follow-up study in Finland. Int J Epidemiol 2015;44:108–116. DOI: 10.1093/ije/dyu235. (Cit. on p. 117).

344. Bandosz P, O'Flaherty M, Rutkowski M et al. A victory for statins or a defeat for diet policies? Cholesterol falls in Poland in the past decade: a modeling study. Int J Cardiol 2015;185:313–319. DOI: 10.1016/j.ijcard.2015.03.079 (cit. on p. 117).

345. Bandosz P, O'Flaherty M, Drygas W et al. Decline in mortality from coronary heart disease in Poland after socioeconomic transformation: modelling studyrganisaion. BMJ 2012;344:d8136. DOI: 10.1136/bmj.d8136 (cit. on pp. 117, 132).

346. Shekelle RB, Shryock AM, Paul O et al. Diet, serum cholesterol, and death from coronary heart disease: the Western Electric study. N Engl J Med 1981;304:65–70. DOI: 10.1056/NEJM198101083040201. (Cit. on p. 117).

347. Food and Agriculture Organisation of the United Nations. FAOSTAT. 2014. Available at: http://faostat.fao.org/default.aspx (accessed on 16th June 2014) (cit. on p. 117).

348. CG181 Lipid modification (update): costing report. National Institute for Health and Care Excellence, 2014. Available at: http://www.nice.org.uk/guidance/cg181/resources/cg181-lipid-modification-update-costing-report (accessed on 27th Sept. 2014) (cit. on p. 117).

349. Ellis JJ, Erickson SR, Stevenson JG, Bemstein SJ, Stiles RA and Fendrick MA. Suboptimal statin adherence and discontinuation in primary and secondary prevention populations. J Gen Intern Med 2004;19:638–645. DOI: 10.1111/j.1525-1497.2004.30516.x. (Cit. on p. 118).

350. Chaudhry HJ and McDermott B. Recognizing and improving patient nonadherence to statin therapy. Curr Atheroscler Rep 2008;10:19–24. DOI: 10.1007/s11883-008-0004-4. (Cit. on p. 118).

351. Bouchard MH, Dragomir A, Blais L, Bérard A, Pilon D and Perreault S. Impact of adherence to statins on coronary artery disease in primary prevention. Br J Clin Pharmacol 2007;63:698–708. DOI: 10.1111/j.1365-2125.2006.02828.x. (Cit. on p. 118).

352. Wallach-Kildemoes H, Andersen M, Diderichsen F and Lange T. Adherence to preventive statin therapy according to socioeconomic position. Eur J Clin Pharmacol 2013;69:1553–1563. DOI: 10.1007/s00228-013-1488-6 (cit. on pp. 118, 123, 132, 194).

353. Sugiyama T, Tsugawa Y, Tseng CH, Kobayashi Y and Shapiro MF. Different time trends of caloric and fat intake between statin users and nonusers among US adults: gluttony in the time of statins? JAMA Internal Medicine 2014;174:1038–1045. DOI: 10.1001/jamainternmed.2014.1927 (cit. on p. 118).

354. Jørgensen T, Capewell S, Prescott E et al. Population-level changes to promote cardiovascular health. Eur J Prev Cardiol 2013;20:409–421. DOI: 10.1177/2047487312441726. (Cit. on p. 118).

355. Brunner EJ, Marmot MG, White IR et al. Gender and employment grade differences in blood cholesterol, apolipoproteins and haemostatic factors in the Whitehall II study. Atherosclerosis 1993;102:195–207. DOI: 10.1016/0021-9150(93)90162-N. (Cit. on p. 119).

356. Gazmararian JA, Kripalani S, Miller MJ, Echt KV, Ren J and Rask K. Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. J Gen Intern Med 2006;21:1215–1221. DOI: 10.1111/j.1525-1497.2006.00591.x. (Cit. on pp. 119, 167).

357. Hippisley-Cox J, Parker C, Coupland C and Vinogradova Y. Inequalities in the primary care of patients with coronary heart disease and serious mental health problems: a cross-sectional study. Heart 2007;93:1256–1262. DOI: 10.1136/hrt.2006.110171. (Cit. on pp. 119, 167).

358. Hinchliffe A. Patient adherence to treatment with statins for the prevention of cardiovascular disease. Cardiff: Public Health Wales NHS Trust, 2011. Available at: http://www2.nphs.wales.nhs.uk:8080/pharmaceuticalphtdocs.nsf/1f8687d8da97650980256fa30051b0be/f552974f8741744f8025786300351952d/$FILE/Adherence%20with%20statins%20to%20prevent%20CVD-%20literature%20review%202011%20v1%200.pdf (accessed on 16th Oct. 2016) (cit. on p. 119).

359. NHS Choices. NHS health check. Available at: http://www.healthcheck.nhs.uk/ (accessed on 23rd Jan. 2016) (cit. on p. 121).

360. Capewell S, McCartney M and Holland W. Invited debate: NHS Health Checks—a naked emperor? J Public Health 2015;37:187–192. DOI: 10.1093/pubmed/fdv063. (Cit. on p. 121).

361. Waterall J, Greaves F, Kearney M and Fenton KA. Invited debate: NHS Health Check: an innovative component of local adult health improvement and well-being programmes in England. J Public Health 2015;37:177–184. DOI: 10.1093/pubmed/fdv062. (Cit. on p. 121).

362. Braillon A, Bewley S, Pisinger C, Fisken RA and Richmond C. NHS health checks are a waste of resources. BMJ feb24 10 2015;350:h1006. DOI: 10.1136/bmj.h1006. (Cit. on p. 121).

363. O'Dowd A. MPs call for review of NHS health checks. BMJ oct28 2 2014;349:g6498–g6498. DOI: 10.1136/bmj.g6498. (Cit. on pp. 121, 123).

364. Department of Health. Government response to the House of Commons Science and Technology Committee report on national health screening. 2015. Available at: https: //www.gov.uk/government/publications/national-health-screening-government-response (accessed on 7th Apr. 2015) (cit. on p. 121).

365. He FJ, Brinsden HC and MacGregor GA. Salt reduction in the United Kingdom: a successful experiment in public health. J Hum Hypertens 2014;28:345–352. DOI: 10. 1038/jhh.2013.105. (Cit. on pp. 121, 145).

366. Levy DT, Currie L and Clancy L. Tobacco control policy in the UK: blueprint for the rest of Europe? Eur J Public Health 2013;23:201–206. DOI: 10.1093/eurpub/cks090 (cit. on p. 121).

367. Krogsboll LT, Jørgensen KJ, Gronhoj Larsen C and Gotzsche PC. General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. BMJ nov20 3 2012;345:e7191. DOI: 10.1136/bmj. e7191. (Cit. on p. 121).

368. Jørgensen T, Jacobsen RK, Toft U, Aadahl M, Glumer C and Pisinger C. Effect of screening and lifestyle counselling on incidence of ischaemic heart disease in general population: Inter99 randomised trial. BMJ jun09 2 2014;348:g3617. DOI: 10.1136/ bmj.g3617. (Cit. on p. 121).

369. Manuel DG, Kwong K, Tanuseputro P et al. Effectiveness and efficiency of different guidelines on statin treatment for preventing deaths from coronary heart disease: modelling study. BMJ 2006;332:1419. DOI: 10.1136/bmj.38849.487546.DE. (Cit. on p. 121).

370. Chang KCM, Soljak M, Lee JT et al. Coverage of a national cardiovascular risk assessment and management programme (NHS Health Check): Retrospective database study. Prev Med 2015;78:1–8. DOI: 10.1016/j.ypmed.2015.05.022 (cit. on pp. 122, 123, 194).

371. Robson J, Dostal I, Sheikh A et al. The NHS Health Check in England: an evaluation of the first 4 years. BMJ Open 2016;6:e008840. DOI: 10.1136/bmjopen-2015-008840. (Cit. on p. 122).

372. Gidlow C, Ellis N, Randall J et al. Method of invitation and geographical proximity as predictors of NHS Health Check uptake. J Public Health 2015;37:195–201. DOI: 10.1093/pubmed/fdu092 (cit. on p. 122).

373. Attwood S, Morton K and Sutton S. Exploring equity in uptake of the NHS Health Check and a nested physical activity intervention trial. J Public Health 2015:fdv070. DOI: 10.1093/pubmed/fdv070. (Cit. on p. 122).

374. Baker C, Loughren EA, Crone D and Kallfa N. A process evaluation of the NHS Health Check care pathway in a primary care setting. J Public Health 2015;37:202–209. DOI: 10.1093/pubmed/fdv053. (Cit. on p. 122).

375. Public Health England. Explore NHS health check data. 2015. Available at: http://www.healthcheck.nhs.uk/interactive_map/ (accessed on 25th Feb. 2015) (cit. on p. 123).

376. Forster AS, Dodhia H, Booth H et al. Estimating the yield of NHS Health Checks in England: a population-based cohort study. J Public Health 2015;37:234–40. DOI: 10.1093/pubmed/fdu079. (Cit. on pp. 123, 247, 248).

377. Hutfless S, Gudzune KA, Maruthur N et al. Strategies to prevent weight gain in adults: a systematic review. Am J Prev Med 2013;45:e41–e51. DOI: 10.1016/j.amepre.2013.07.013. (Cit. on p. 123).

378. Bauld L, Bell K, McCullough L, Richardson L and Greaves L. The effectiveness of NHS smoking cessation services: a systematic review. J Public Health 2010;32:71–82. DOI: 10.1093/pubmed/fdp074. (Cit. on pp. 123, 194).

379. West R and Owen L. Estimates of 52-week continuous abstinence rates following selected smoking cessation interventions in England. Smoking in England, 2012. Available at: http://www.smokinginengland.info/downloadfile/?type=report&src=5 (cit. on pp. 123, 194).

380. Sharma A, Hauck K, Hollingsworth B and Siciliani L. The effects of taxing sugar-sweetened beverages across different income groups. Health Econ 2014;23:1159–1184. DOI: 10.1002/hec.3070 (cit. on pp. 123, 195).

381. Briggs ADM, Mytton OT, Kehlbacher A, Tiffin R, Rayner M and Scarborough P. Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in UK: econometric and comparative risk assessment modelling study. BMJ 2013;347:f6189. DOI: 10.1136/bmj.f6189 (cit. on pp. 123, 195).

382. Cabrera Escobar MA, Veerman JL, Tollman SM, Bertram MY and Hofman KJ. Evidence that a tax on sugar sweetened beverages reduces the obesity rate: a meta-analysis. BMC Public Health 2013;13:1072. DOI: 10.1186/1471-2458-13-1072 (cit. on pp. 123, 195).

383. Webster JL, Dunford EK, Hawkes C and Neal BC. Salt reduction initiatives around the world: J Hypertens 2011;29:1043–1050. DOI: 10.1097/HJH.0b013e328345ed83. (Cit. on pp. 123, 135).

384. Bartlett S, Klerman J, Olsha L et al. Evaluation of the Healthy Incentives Pilot (HIP) final report. U.S. Department of Agriculture, Food and Nutrition Service, 2014. Available at: http://www.fns.usda.gov/sites/default/files/HIP-Final.pdf (cit. on pp. 124, 195).

385. Nnoaham KE, Sacks G, Rayner M, Mytton O and Gray A. Modelling income group differences in the health and economic impacts of targeted food taxes and subsidies. Int J Epidemiol 2009;38:1324–1333. DOI: 10.1093/ije/dyp214. (Cit. on pp. 124, 133, 195).

386. Allen K, Gillespie DOS, Guzman-Castillo M, Diggle PJ, Capewell S and O'Flaherty M. Future trends and inequalities in premature coronary deaths in England: modelling study. Int J Cardiol 2016;203:290–297. DOI: 10.1016/j.ijcard.2015.10.077. (Cit. on p. 131).

387. Schuetz CA, Alperin P, Guda S et al. A standardized vascular disease health check in europe: a cost-effectiveness analysis. PLOS One 2013;8. Ed. by Federici M:e66454. DOI: 10.1371/journal.pone.0066454. (Cit. on p. 132).

388. Hobbs FR, Banach M, Mikhailidis DP, Malhotra A and Capewell S. Is statin-modified reduction in lipids the most important preventive therapy for cardiovascular disease? A pro/con debate. BMC Med 2016;14:4. DOI: 10.1186/s12916-016-0550-5. (Cit. on p. 132).

389. Wallach-Kildemoes H, Diderichsen F, Krasnik A, Lange T and Andersen M. Is the high-risk strategy to prevent cardiovascular disease equitable? A pharmacoepidemiological cohort study. BMC Public Health 2012;12:610. DOI: 10.1186/1471-2458-12-610 (cit. on p. 132).

390. Bender AM, Kawachi I, Jørgensen T and Pisinger C. Neighborhood deprivation is strongly associated with participation in a population-based health check. PLOS One 2015;10. Ed. by Zeeb H:e0129819. DOI: 10.1371/journal.pone.0129819 (cit. on p. 132).

391. Franco M, Bilal U, Orduñez P et al. Population-wide weight loss and regain in relation to diabetes burden and cardiovascular mortality in Cuba 1980-2010: repeated cross sectional surveys and ecological comparison of secular trends. BMJ 2013;346:f1515–f1515. DOI: 10.1136/bmj.f1515. (Cit. on p. 132).

392. Danaei G, Finucane MM, Lin JK et al. National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5.4 million participants. Lancet 2011;377:568–577. DOI: 10.1016/S0140-6736(10)62036-3. (Cit. on p. 132).

393. Molarius A, Parsons RW, Dobson AJ et al. Trends in cigarette smoking in 36 populations from the early 1980s to the mid-1990s: findings from the WHO MONICA Project. Am J Public Health 2001;91:206–212. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1446542/ (accessed on 10th Apr. 2015) (cit. on p. 132).

394. Vos T, Carter R, Barendregt J et al. Assessing cost-effectiveness in prevention. Melbourne: The University of Queensland, Brisbane, and Deakin University, 2010. Available at: http://www.sph.uq.edu.au/docs/BODCE/ACE-P/ACE-Prevention_final_report.pdf (cit. on p. 133).

395. Lloyd-Williams F, Bromley H, Orton L et al. Smorgasbord or symphony? Assessing public health nutrition policies across 30 European countries using a novel framework. BMC Public Health 2014;14:1195. DOI: 10.1186/1471-2458-14-1195. (Cit. on p. 133).

396. Bødker M, Pisinger C, Toft U and Jørgensen T. The rise and fall of the world's first fat tax. Health Policy 2015;119:737–42. DOI: 10.1016/j.healthpol.2015.03.003. (Cit. on p. 133).

397. Leifert RM and Lucinda CR. Linear symmetric "fat taxes": evidence from Brazil. Appl Econ Perspect Policy 2015;37:634–666. DOI: 10.1093/aepp/ppu062. (Cit. on p. 133).

398. Madden D. The poverty effects of a 'fat-tax' in Ireland. Health Econ 2015;24:104–121. DOI: 10.1002/hec.3006. (Cit. on p. 133).

399. Trieu K, Neal B, Hawkes C et al. Salt reduction initiatives around the world – a systematic review of progress towards the global target. PLOS One 2015;10. Ed. by DeAngelis MM:e0130247. DOI: 10.1371/journal.pone.0130247. (Cit. on pp. 135, 136).

400. Wyness LA, Butriss JL and Stanner SA. Reducing the population's sodium intake: the UK Food Standards Agency's salt reduction programme. Public Health Nutr 2012;15:254–261. DOI: 10.1017/S1368980011000966. (Cit. on pp. 135, 145).

401. Cobiac LJ, Vos T and Veerman JL. Cost-effectiveness of interventions to reduce dietary salt intake. Heart 2010;96:1920–1925. DOI: 10.1136/hrt.2010.199240. (Cit. on p. 135).

402. Millett C, Laverty AA, Stylianou N, Bibbins-Domingo K and Pape UJ. Impacts of a national strategy to reduce population salt intake in england: serial cross sectional study. PLOS One 2012;7. Ed. by Bochud M:e29836. DOI: 10.1371/journal.pone.0029836. (Cit. on p. 144).

403. Dodhia H, Phillips K, Zannou MI, Airoldi M and Bevan G. Modelling the impact on avoidable cardiovascular disease burden and costs of interventions to lower SBP in the England population: J Hypertens 2012;30:217–226. DOI: 10.1097/HJH.0b013e32834d86ee. (Cit. on p. 144).

404. Hendriksen MAH, van Raaij JMA, Geleijnse JM, Breda J and Boshuizen HC. Health gain by salt reduction in Europe: a modelling study. PLOS One 2015;10. DOI: 10.1371/journal.pone.0118873. (Cit. on p. 144).

405. Knai C, Petticrew M, Durand MA et al. Has a public–private partnership resulted in action on healthier diets in England? An analysis of the Public Health Responsibility Deal food pledges. Food Policy 2015;54:1–10. DOI: 10.1016/j.foodpol.2015.04.002. (Cit. on p. 145).

406. Trinquart L, Johns DM and Galea S. Why do we think we know what we know? A metaknowledge analysis of the salt controversy. Int J Epidemiol 2016;45:251–260. DOI: 10.1093/ije/dyv184. (Cit. on p. 145).

407. Stolarz-Skrzypek K, Kuznetsova T, Thijs L and al et. Fatal and nonfatal outcomes, incidence of hypertension, and blood pressure changes in relation to urinary sodium excretion. JAMA 2011;305:1777–1785. DOI: 10.1001/jama.2011.574. (Cit. on p. 145).

408. Action on Smoking and Health (ASH). Smoking statistics: illness and death. 2014. Available at: http://ash.org.uk/files/documents/ASH_107.pdf (accessed on 8th Mar. 2015) (cit. on p. 147).

409. Office for National Statistics. Opinions and lifestyle survey, adult smoking habits in Great Britain, 2013. Office for National Statistics. 2014. Available at: http://www.ons.gov.uk/ons/rel/ghs/opinions-and-lifestyle-survey/adult-smoking-habits-in-great-britain--2013/index.html (accessed on 3rd Aug. 2015) (cit. on p. 147).

410. Hiscock R, Bauld L, Amos A and Platt S. Smoking and socioeconomic status in England: the rise of the never smoker and the disadvantaged smoker. J Public Health 2012;34:390–396. DOI: 10.1093/pubmed/fds012 (cit. on p. 147).

411. Office for National Statistics. Do smoking rates vary between more and less advantaged areas? 2014. Available at: http://www.ons.gov.uk/ons/rel/disability-and-health-measurement/do-smoking-rates-vary-between-more-and-less-advantaged-areas-/2012/sty-smoking-rates.html (accessed on 3rd Aug. 2015) (cit. on p. 147).

412. Giesinger I, Goldblatt P, Howden-Chapman P, Marmot M, Kuh D and Brunner E. Association of socioeconomic position with smoking and mortality: the contribution of early life circumstances in the 1946 birth cohort. J Epidemiol Community Health 2014;68:275–279. DOI: 10.1136/jech-2013-203159 (cit. on p. 147).

413. Joossens L and Raw M. The tobacco control scale 2013 in Europe. In: *Sixth European Conference on Tobacco or Health-ECToH, Istanbul, Turkey.* 2014. Available at: http://www.europeancancerleagues.org/images/TobaccoControl/TCS_2013_in_Europe_13-03-14_final_1.pdf (accessed on 3rd Aug. 2015) (cit. on p. 147).

414. World Health Organisation. WHO report on the global tobacco epidemic, 2013: enforcing bans on tobacco advertising, promotion and sponsorship. World Health Organisation, 2013. Available at: http://www.who.int/iris/bitstream/10665/85380/1/9789241505871_eng.pdf (cit. on p. 147).

415. Allen K, Kypridemos C, Hyseni L et al. The effects of maximising the UK's tobacco control score on inequalities in smoking prevalence and premature coronary heart disease mortality: a modelling study. BMC Public Health 2016;16:292. DOI: 10.1186/s12889-016-2962-8. (Cit. on pp. 148, 160).

416. Warner KE. An endgame for tobacco? Tob Control 2013;22:i3–5. DOI: 10.1136/tobaccocontrol-2013-050989 (cit. on p. 148).

417. McDaniel PA, Smith EA and Malone RE. The tobacco endgame: a qualitative review and synthesis. Tob Control 2015;25:594–604. DOI: 10.1136/tobaccocontrol-2015-052356. (Cit. on p. 148).

418. Khoo D, Chiam Y, Ng P, Berrick AJ and Koong HN. Phasing-out tobacco: proposal to deny access to tobacco for those born from 2000. Tob Control 2010;19:355–360. DOI: 10.1136/tc.2009.031153 (cit. on p. 148).

419. Berrick AJ. The tobacco-free generation proposal. Tob Control 2013;22:i22–i26. DOI: 10.1136/tobaccocontrol-2012-050865 (cit. on p. 148).

420. Proctor RN. Why ban the sale of cigarettes? The case for abolition. Tob Control 2013;22:i27–i30. DOI: 10.1136/tobaccocontrol-2012-050811 (cit. on pp. 148, 159).

421. Daynard RA. Doing the unthinkable (and saving millions of lives). Tob Control 2009;18:2–3. DOI: 10.1136/tc.2008.028308 (cit. on p. 148).

422. Waldhuter L. Tasmania 'could have strictest smoking laws in the world'. 2015. Available at: http://www.abc.net.au/news/2015-12-21/state-plan-to-lift-tasmanias-legal-smoking-age-above-18/7044622 (accessed on 30th June 2016) (cit. on p. 148).

423. British Medical Association. BMA - Doctors back cigarette ban to those born after 2000. 2016. Available at: https://www.bma.org.uk/news/2014/june/doctors-back-cigarette-ban-to-those-born-after-2000 (accessed on 29th June 2016) (cit. on p. 148).

424. Givel M. Tobacco policymaking and administration in Bhutan from 1991 to 2009. International Journal of Public Administration 2011;34:775–782. DOI: 10.1080/01900692.2011.597479 (cit. on pp. 148, 159).

425. Bagcchi S. Turkmenistan bans tobacco sales. Lancet Resp Med 2016;4:180. DOI: 10.1016/s2213-2600(16)00047-3 (cit. on p. 148).

426. Givel MS. Tobacco use policymaking and administration in Bhutan. 2009. Available at: https://works.bepress.com/michael_givel/19/download/ (accessed on 23rd June 2016) (cit. on pp. 149, 159).

427. Bhimjiyani A, Knuchel-Takano A and Hunt D. Aiming high: why the UK should aim to be tobacco-free. Research rep. UK: UK Health Forum & Cancer Research UK, 2016. Available at: http://nhfshare.heartforum.org.uk/RMAssets/Modelling/CRUKSmoking/Tobacco%20Report%20FINAL%20PDF.pdf (cit. on pp. 158, 160).

428. Isett KR. In and across bureaucracy: structural and administrative issues for the tobacco endgame. Tob Control suppl 1 2013;22:i58–i60. DOI: 10.1136/tobaccocontrol-2012-050828. (Cit. on p. 159).

429. Chapman S and Wakefield MA. Large-scale unassisted smoking cessation over 50 years: lessons from history for endgame planning in tobacco control. Tob Control 2013;22 Suppl 1:i33–35. DOI: 10.1136/tobaccocontrol-2012-050767 (cit. on p. 159).

430. Fernández E, Lugo A, Clancy L, Matsuo K, La Vecchia C and Gallus S. Smoking dependence in 18 European countries: hard to maintain the hardening hypothesis. Prev Med 2015;81:314–319. DOI: 10.1016/j.ypmed.2015.09.023. (Cit. on p. 159).

431. Capewell S and Lilford R. Are nanny states healthier states? BMJ 2016;355. DOI: 10.1136/bmj.i6341. (Cit. on p. 159).

432. Nutt DJ, King LA and Phillips LD. Drug harms in the UK: a multicriteria decision analysis. Lancet 2010;376:1558–1565. DOI: 10.1016/S0140-6736(10)61462-6. (Cit. on p. 159).

433. Gallus S, Lugo A, Fernandez E et al. Support for a tobacco endgame strategy in 18 European countries. Prev Med 2014;67:255–258. DOI: 10.1016/j.ypmed.2014.08.001. (Cit. on p. 159).

434. Lykke M, Pisinger C and Glümer C. Ready for a goodbye to tobacco? — Assessment of support for endgame strategies on smoking among adults in a Danish regional health survey. Prev Med 2016;83:5–10. DOI: 10.1016/j.ypmed.2015.11.016. (Cit. on p. 159).

435. Shahab L and West R. Public support in England for a total ban on the sale of tobacco products. Tob Control 2016;19:143–147. DOI: 10.1136/tc.2009.033415 (cit. on p. 159).

436. Malone R, McDaniel P and Smith E. It is time to plan the tobacco endgame. BMJ 2014;348:g1453. DOI: 10.1136/bmj.g1453. (Cit. on pp. 159, 160).

437. Malone RE, McDaniel PA and Smith EA. Tobacco control endgames: global initiatives and implications for the UK. Cancer Research UK, 2015 (cit. on p. 160).

438. Cobiac LJ, Ikeda T, Nghiem N, Blakely T and Wilson N. Modelling the implications of regular increases in tobacco taxation in the tobacco endgame. Tob Control e2 2015;24:e154–e160. DOI: 10.1136/tobaccocontrol-2014-051543. (Cit. on p. 160).

439. Briggs ADM, Mizdrak A and Scarborough P. A Statin a Day Keeps the Doctor Away: Comparative Proverb Assessment Modelling Study. BMJ 2013;347:f7267. DOI: 10.11 36/bmj.f7267. (Cit. on p. 164).

440. Nichols M, Scarborough P, Allender S and Rayner M. What is the optimal level of population alcohol consumption for chronic disease prevention in England? Modelling the impact of changes in average consumption levels. BMJ Open 2012;2:e000957. DOI: 10.1136/bmjopen-2012-000957. (Cit. on p. 164).

441. Cobiac LJ, Tam K, Veerman L and Blakely T. Taxes and subsidies for improving diet and population health in Australia: a cost-effectiveness modelling study. PLOS Medicine 2017;14:e1002232. DOI: 10.1371/journal.pmed.1002232. (Cit. on p. 165).

442. Mytton OT, Tainio M, Ogilvie D, Panter J, Cobiac L and Woodcock J. The modelled impact of increases in physical activity: the effect of both increased survival and reduced incidence of disease. European Journal of Epidemiology 2017:1–16. DOI: 10.1007/s10654-017-0235-1. (Cit. on p. 166).

443. Harding A and Harding A. Microsimulation and public policy. Amsterdam : New York: Emerald Group Publishing Limited, 1996. 546 pp. (cit. on p. 168).

444. Li J and O'Donoghue C. A survey of dynamic microsimulation models: uses, model structure and methodology. International Journal of Microsimulation 2013;6:3–55. Available at: http://merit.unu.edu/publications/wppdf/2012/wp2012-002.pdf (accessed on 21st July 2014) (cit. on p. 168).

445. Lay-Yee R and Cotterell G. The role of microsimulation in the development of public policy. In: *Policy Practice and Digital Science*. Ed. by Janssen M, Wimmer MA and Deljoo A. Public Administration and Information Technology 10. Springer International Publishing, 2015:305–320. DOI: 10.1007/978-3-319-12784-2_14. (Cit. on p. 168).

446. Abraham JM. Using microsimulation models to inform U.S. health policy making. Health Serv Res 2013;48:686–695. DOI: 10.1111/1475-6773.12052. (Cit. on p. 168).

447. Ringel JS, Eibner C, Girosi F, Cordova A and McGlynn EA. Modeling health care policy alternatives. Health Serv Res 2010;45:1541–1558. DOI: 10.1111/j.1475-6773. 2010.01146.x. (Cit. on p. 168).

448. Ballard-Barbash R, Hunsberger S, Alciati MH et al. Physical activity, weight control, and breast cancer risk and survival: clinical trial rationale and design considerations. J Natl Cancer Inst 2009;101:630–643. DOI: 10.1093/jnci/djp068. (Cit. on p. 168).

449. O'Luanaigh C. New results indicate that new particle is a Higgs boson | CERN. http://home.cern/about/updates/2013/03/new-results-indicate-new-particle-higgs-boson. 2013. (Accessed on 25th Oct. 2016) (cit. on p. 169).

450. Fox J, Dangler C, Taylor N, King A, Koh T and Wang T. High-salt diet induces gastric epithelial hyperplasia and parietal cell loss, and enhances Helicobacter pylori colonization in C57BL/6 mice. Cancer Res 1999;59. bibtex: fox_high-salt_1999:4823–4828 (cit. on p. 169).

451. Tatematsu M, Takahashi M, Fukushima S, Hananouchi M and Shirai T. Effects in rats of sodium chloride on experimental gastric cancers induced by N-methyl-N-nitro-N-nitrosoguanidine or 4-nitroquinoline-1-oxide. J Natl Cancer Inst 1975;55. bibtex: tatematsu_effects_1975:101–106 (cit. on p. 169).

452. Hill AL, Rand DG, Nowak MA and Christakis NA. Infectious disease modeling of social contagion in networks. PLOS Comput Biol 2010;6:e1000968. DOI: 10.1371/journal.pcbi.1000968. (Cit. on p. 170).

453. Bishai D, Paina L, Li Q, Peters DH and Hyder AA. Advancing the application of systems thinking in health: why cure crowds out prevention. Health Res Policy Syst 2014;12. DOI: 10.1186/1478-4505-12-28. (Cit. on p. 171).

454. Kypridemos C, Bandosz P, Hickey GL et al. Quantifying the contribution of statins to the decline in population mean cholesterol by socioeconomic group in England 1991 - 2012: a modelling study. PLOS One 2015;10:e0123112. DOI: 10.1371/journal.pone.0123112. (Cit. on p. 172).

455. Office for National Statistics. Migration statistics quarterly report: August 2016. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/migrationstatisticsquarterlyreport/august2016. 2016. (Accessed on 11th Oct. 2016) (cit. on p. 173).

456. The Migration Observatory, University of Oxford. Health of migrants in the UK: what do we know? http://www.migrationobservatory.ox.ac.uk/resources/briefings/health-of-migrants-in-the-uk-what-do-we-know/. 2016. (Accessed on 11th Oct. 2016) (cit. on p. 173).

457. Scholes S, Coombs N, Pedisic Z et al. Age- and sex-specific criterion validity of the health survey for England physical activity and sedentary behavior assessment questionnaire as compared with accelerometry. Am J Epidemiol 2014;179:1493–1502. DOI: 10.1093/aje/kwu087. (Cit. on p. 174).

458. Ford ES, Ajani UA, Croft JB et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med 2007;356:2388–2398. DOI: 10 . 1056 / NEJMsa053935 (cit. on p. 174).

459. Kannel WB, Neaton JD, Wentworth D et al. Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for the MRFIT. Am Heart J 1986;112:825–836. DOI: 10.1016/0002-8703(86)90481-3. (Cit. on p. 175).

460. Critchley JA and Capewell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. JAMA 2003;290:86–97. DOI: 10.1001/jama.290.1.86. (Cit. on p. 175).

461. Obermeyer Z and Emanuel EJ. Predicting the future — Big data, machine learning, and clinical medicine. 2016;375:1216–1219. DOI: 10.1056/NEJMp1606181. (Cit. on p. 177).

462. van Staa TP, Goldacre B, Buchan I and Smeeth L. Big health data: the need to earn public trust. BMJ 2016;354:i3636. DOI: 10.1136/bmj.i3636. (Cit. on p. 177).

463. Su TL, Jaki T, Hickey GL, Buchan I and Sperrin M. A review of statistical updating methods for clinical prediction models. Stat Methods Med Res 2016. in press. DOI: 10.1177/0962280215626466 (cit. on p. 177).

464. Buchan I, Winn J and Bishop C. A unified modeling approach to data-intensive healthcare. In: *The fourth paradigm: data-intensive scientific discovery.* Microsoft Research, 2009. Available at: https://www.microsoft.com/en-us/research/publication/a-unified-modeling-approach-to-data-intensive-healthcare/ (accessed on 21st Oct. 2016) (cit. on p. 177).

465. Khoury MJ, Iademarco MF and Riley WT. Precision public health for the era of precision medicine. Am J Prev Med 2016;50:398–401. DOI: 10.1016/j.amepre.2015.08.031. (Cit. on p. 177).

466. Yadav M, Malhotra P, Vig L, Sriram K and Shroff G. ODE - Augmented training improves anomaly detection in sensor data from machines. arXiv preprint arXiv:1605.01534 2016. Available at: http://arxiv.org/abs/1605.01534 (accessed on 13th Oct. 2016) (cit. on p. 178).

467. Voltaire. Dictionnaire philosophique. 1764 (cit. on p. 178).

468. Belgorodski N, Greiner M, Tolksdorf K and Schueller K. rriskDistributions:fitting distributions to given data or known quantiles. Version 2.1. 2015. Available at: http://CRAN.R-project.org/package=rriskDistributions (cit. on p. 183).

469. Forster AS, Burgess C, Dodhia H et al. Do health checks improve risk factor detection in primary care? Matched cohort study using electronic health records. J Public Health 2015:fdv119. DOI: 10.1093/pubmed/fdv119. (Cit. on pp. 194, 195).

470. Collins M, Mason H, O'Flaherty M, Guzman-Castillo M, Critchley J and Capewell S. An economic evaluation of salt reduction policies to reduce coronary heart disease in England: a policy modeling study. Value Health 2014;17:517–524. DOI: 10.1016/j.jval.2014.03.1722 (cit. on p. 195).

471.   Symmons D, Turner G, Webb R et al. The prevalence of rheumatoid arthritis in the
        United Kingdom: new estimates for a new century. Rheumatology 2002;41:793–800.
        DOI: 10.1093/rheumatology/41.7.793. (Cit. on p. 247).

472.   Usher-Smith JA, Pritchard J, Poole S and Griffin SJ. Offering statins to a population
        attending health checks with a 10-year cardiovascular disease risk between 10% and
        20%. Int J Clin Pract 2015;69. DOI: 10.1111/ijcp.12742. (Cit. on p. 248).