The University Of Sheffield.

Longitudinal clustering of health risk behaviours and their association with multimorbidity in the United Kingdom

Alisha Suhag

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
School of Medicine and Population Health
April 2024

Supervisors:

Professor John Holmes
University of Sheffield

Professor Thomas L Webb
University of Sheffield

Acknowledgments

This DPhil would not have been possible without some incredible people who have supported me.

First, I would like to thank my supervisors Professor John Holmes and Professor Thomas L. Webb. Working with and learning from you has been one of the best parts of this PhD for me. You have both set the bar very high for what dependability and consistency as a professional looks like. Thank you for your expertise, time, and patience in shaping me as a researcher.

I would like to thank the Healthy Lifespan Institute for funding my PhD and creating such a collaborative network to work along.

Thank you to my Mum, for encouraging me in everything. Your courage to dream opened up so many possibilities for us.

Last but foremost, thank you to my partner Saud. You inspire me. Thank you for your love and strength. You make everything fun.

I dedicate this work to my beloved late father, Jaber Singh Suhag. He made it possible for me to be here, to do this.

Abstract

Background

Understanding the relationship between modifiable risk behaviours such as smoking, poor nutrition, alcohol consumption, and physical inactivity (termed 'SNAP' behaviours) and multimorbidity is crucial for disease prevention. Risk behaviours often cluster in specific combinations within distinct subpopulations, but little is known about how this clustering changes with age, and how it is associated with multimorbidity.

Aims

Thus, the present research aimed to: i) explore whether and how the SNAP behaviours cluster over time, ii) investigate whether and how membership in different behavioural clusters varies by socio-demographic characteristics, and iii) examine which, if any, behavioural clusters are prospectively associated with multimorbidity over time.

Methods

Using data from longitudinal surveys, a first study analysed data from the English Longitudinal Study of Ageing (ELSA) on older adults (aged 50+). A second study replicated the first using data from the UK Household Longitudinal Study (UKHLS) on a broader age group of adults (aged 16+). A repeated latent class analysis was conducted to identify clusters of SNAP behaviours. Logistic regressions were used to examine how the clusters were associated with socio-demographic characteristics and disease status. Further, a third study conducted latent class moderation to examine whether socio-demographic characteristics moderated the relationship between the clusters identified using the ELSA dataset and disease status.

Results

Seven clusters of individuals with distinct SNAP behaviour patterns and sociodemographic profiles were identified in older adults and the general adult sample. Behaviour patterns within these clusters remained fairly stable over time. Across both studies, broad similarities were found in the behaviour profiles and sociodemographic characteristics of clusters that had both the highest and the lowest prevalence of multimorbidity. Clusters characterised by sociodemographic profiles with significant socio-economic disadvantages had a higher prevalence of multimorbidity and other health conditions–despite engaging in fewer risky behaviours than some other clusters. The moderation analysis did not reveal any significant interactions.

Conclusions

A strong social gradient was observed in the relationship between risk behaviour clusters and health outcomes such that the clusters characterised by sociodemographic profiles with significant social disadvantages had worse health outcomes–despite engaging in fewer risky behaviours.

Publications

Publications contributing to DPhil thesis

Suhag, A., Webb, T. L., & Holmes, J. (2024). Longitudinal clustering of health behaviours and their association with multimorbidity in older adults in England: A latent class analysis. *Plos one, 19*(1), e0297422.

Suhag, A., Webb, T. L., & Holmes, J. (2024). Moderating effects of socio-demographic characteristics on the relationship between risk behaviour clusters and multimorbidity in UK older adults. (*To be submitted to Plos one*)

Suhag, A., Webb, T. L., & Holmes, J. (2024). Longitudinal clustering of health behaviours and their association with multimorbidity: Evidence from Understanding Society (UKHLS) (*To be submitted to the International Journal of Epidemiology*)

Research achievements 2020-2024

Publications

Suhag A., Rerkasem A., Kulprachakarn K., Parklak W., Angkurawaranon C., Rerkasem K. & Derraik J.G.B (2022) Long-term health associated with small and large for gestational age births among young Thai adults. *Children, 9*(6).

Submitted manuscripts

Glanz, K., Holman, D., **Suhag A.** Association of health behavior theory use and intervention effectiveness: An umbrella review. (Submitted to *Health Psychology*)

De-Ville S., Pott S., Wood C., Peng Z., Farooq M.Z., Blouchou P., **Suhag A.,** Murphy E., Hathway E.A. Improving hospitality venue ventilation via behavioural change as a response to the COVID- 19 pandemic. (Submitted to *Building and Environment*)

Commissioned Report

Glanz K., **Suhag A.,** Holman D. (2023). Review of the use of behavioural science in health promotion and in health promotion implementation strategies, plans and interventions: a gap analysis. (Under review - Commissioned by World Health Organisation Behavioural Insights Unit)

Published Conference Abstracts

Suhag, A., Webb, T. L., & Holmes, J. (2023). Longitudinal clustering of risk behaviours and their association with multimorbidity in older adults. European Journal of Public Health, 33(Supplement_2), ckad160-363.

Suhag, A., Holmes, J., & Webb, T. L. (2023). Longitudinal clustering of health behaviours and their association with health outcomes in older adults in England: a latent class analysis. Population Medicine, 5 (Supplement).

Seminars and Workshops

- CLOSER Emerging Longitudinal Scholars webinar. Feb 2024. CLOSER, *UCL Social Research Institute.*
- Creating Finite Mixture Models using MPlus Workshop. Jan 2024. Division of Population Health, *University of Sheffield.*
- Healthy Lifespan Institute PhD Symposium. June 2023. *University of Sheffield.*

Awards

University of Sheffield Postgraduate Research Publication Scholarship (2024)

Table of contents

Chapter 1: Introduction

1.1 Research background

"Multimorbidity cuts across the vertical paradigms in which most health research and policy is envisaged, supported and carried out, reflecting not only specialist interests in particular problems and diseases but also the tendency of research to focus on easily defined issues."

(Mercer et al., 2009; p.1)

Multimorbidity, the presence of two or more conditions in the same individual, poses a significant public health challenge (Head et al., 2021a). It affects a substantial portion of the general population, with studies estimating a prevalence of 33%-42% globally (Nguyen et al., 2019; Ho et al., 2022a). In high-income countries like the United Kingdom (UK), the burden of multimorbidity is projected to rise sharply and stems from various factors (Soley-Bori et al., 2021). These include population ageing, shifts in environmental risks, advancements in healthcare, changes in health literacy, and lifestyle risk behaviours (Xie et al., 2022). There is a pressing need to further understand the burden, risk factors, and consequences of multimorbidity to guide planning and resource allocation for effective prevention, timely intervention, and disease management. To mitigate the burden of multimorbidity, primary and secondary prevention through identifying and targeting shared risk factors for common chronic diseases holds promise.

This thesis investigates the relationship between multimorbidity and four modifiable risk factors—smoking, poor nutrition, alcohol consumption, and physical activity (collectively referred to as SNAP behaviours). SNAP behaviours account for nearly one-third of disability-adjusted life years from chronic conditions (Global Burden of Disease Study Collaborators,

2017). They are the leading causes of the "big four" chronic diseases contributing substantially to the multimorbidity burden—type-2 diabetes, coronary heart disease, chronic obstructive pulmonary disease, and specific cancers (Niebuur et al., 2023). By identifying how SNAP behaviours cluster as people age, this thesis aims to uncover meaningful patterns of risk and protective factors in subgroups of the population at high risk of disease, which can, in turn, facilitate the planning of targeted behavioural interventions.

More specifically, this thesis aimed to advance our understanding of how SNAP risk behaviours cluster over time, the socio-demographic profiles of identified clusters, and whether and how the identified clusters are prospectively associated with the prevalence of multimorbidity and a range of chronic disease outcomes. To do this, I used data from two large, ongoing nationally-representative panel studies in the UK: the English Longitudinal Study of Ageing (ELSA) and the UK Household Longitudinal Study (UKHLS). ELSA focuses on older adults (aged 50 and over), and UKHLS focuses on the general adult population (aged 16 and over). These datasets contain a breadth of behavioural, demographic, and disease outcome variables. The data are longitudinal, providing information on a large population broadly representative of the UK.

1.2 Structure of the thesis

The remainder of this section will summarise each chapter in this thesis.

Chapter 2 provides an overview of the concepts, ideas, and previous research relevant to this thesis. It introduces the concept of multimorbidity, the clustering of risk behaviours over time, and their connection to disease from a life course perspective. The chapter describes

multimorbidity within a research context and further explores the impact and public health implications of rising rates of multimorbidity. It then discusses multimorbidity in the context of the UK, highlighting a need for prevention strategies that focus on modifiable behavioural risk factors as potential targets for intervention. It also discusses the current literature on key risk factors for multimorbidity – particularly smoking, poor nutrition, alcohol consumption, and physical activity (termed SNAP behaviours), highlighting how these risks co-occur and cluster in the population, change with age, and impact disease risk. Finally, following a brief overview of ideas around social determinants of health, the chapter closes by highlighting important gaps in the literature to be addressed in subsequent analytical chapters.

Chapter 3 briefly discusses the rationale behind selecting repeated measures latent class analysis (RMLCA) as the primary methodological approach for identifying clusters of health-risk behaviours over time. This chapter outlines the factors considered in making this choice: the available methods, the longitudinal data, and the levels of missingness in that data. It also discusses alternative methodologies that were considered (such as latent transition analysis) and why they were ruled out in favour of RMLCA.

Chapter 4 is the first analytical study in this thesis and is presented as a published paper (Suhag, Webb and Holmes, 2024). This study examined how SNAP behaviours cluster and change with age in older adults living in England, the socio-demographic profiles of identified clusters, and how the clusters were prospectively associated with multimorbidity. The study identified seven distinct clusters, ranging from a cluster with a low-risk profile across all SNAP behaviours to other clusters engaging in one or more behaviours at riskier levels. Notably, the behaviour patterns within these clusters remained fairly stable over time, indicating minimal changes with age. The study also highlighted socio-economic and demographic disparities: the clusters

characterised by physical inactivity (alongside other risky behaviours) primarily included individuals with lower levels of education and less wealth, whereas the heavy drinking clusters were primarily men. Differences in disease prevalence across clusters were also evident. For instance, *High-risk smokers* had the highest prevalence of respiratory disorders, as anticipated. More unexpectedly, a cluster engaging in only one risky behaviour, the *Abstainer but inactive,* had worse health outcomes than clusters engaging in multiple risky behaviours. Although the *Abstainer but inactive* cluster comprised mostly women and had notable socio-economic disadvantages (i.e. less well-educated and wealthy), a higher prevalence of multimorbidity, complex multimorbidity, and endocrine disorders persisted despite controlling for these socio-demographic characteristics. Overall, this research suggests that the link between risk behaviours and health outcomes is not straightforward but rather complex and highlights the importance of considering social determinants of health alongside risk behaviours for chronic disease prevention.

Chapter 5 extends the analysis in Chapter 4 to a UK-wide longitudinal panel study of adults aged 16 and over, identifying seven clusters with distinct behaviour patterns over time and socio-demographic profiles. Across both studies, behaviour patterns within clusters were largely stable over time. Moreover, the clusters were broadly similar to those identified in Chapter 4. In particular, the behaviour patterns and socio-demographic profiles of clusters with the highest and the lowest prevalence of multimorbidity and most health conditions studied were similar across the two studies. Interestingly, the cluster with the least risky behaviour profile — the *Overall low-risk* cluster — had a higher multimorbidity prevalence than other clusters. The socio-demographic profile of this cluster resembled the cluster with the highest prevalence of multimorbidity in Chapter 4 and predominantly comprised women with socio-economic disadvantages. Similar to the study in Chapter 4, adjusting for socio-demographic factors did not change these results much. This unexpected finding prompted further sensitivity

analyses with age-stratified subsamples, which indicated the same trend. Across both studies, similarities in the socio-demographic and behaviour profiles were also observed for the clusters that were least likely to have multimorbidity or other conditions considered. The chapter further explores potential reasons for and implications of these findings to inform future preventative interventions.

Chapter 6 builds on the study investigated in Chapter 4 by exploring the possibility of socio-demographic characteristics moderating the relationship between risk behaviour clusters and disease outcomes in older adults. It utilises the clusters identified in Chapter 4 as a multidimensional latent predictor to represent subgroups of older adults with similar longitudinal patterns of SNAP behaviours. Consequently, the study examines how socio-demographic characteristics may amplify or attenuate the association between multiple risk behaviours over time and chronic diseases — an investigation that is often neglected as socio-demographic characteristics are typically controlled for in studies investigating the influence of risk behaviours on disease. Though the main effects of this study align with prior research, the absence of significant moderation effects leaves some questions unanswered, potentially due to statistical underpowering.

Chapter 7 synthesises the overall findings of the thesis and contextualises them against the wider literature. It discusses the overarching implications of the research for future research and public health policy, identifying areas for further investigation and potential avenues for intervention, such as including mental health outcomes and disease biomarkers, as well as the need to examine emerging risk behaviours like prolonged sitting, excessive screen time, and inadequate sleep. This chapter also identifies challenges for the future of prevention research. It highlights how the empirical challenges of prevention research are complicated by the

persistent structural and systemic inequities in health and chronic disease. The chapter concludes with a critical reflection on the study design, acknowledging its strengths and limitations, and offers insights into the methodological considerations that shaped our findings and could guide future studies.

Chapter 2: Background

2.1 Chapter Overview

This chapter provides an overview of three core areas that are critical to this thesis. Firstly, the prevalence, costs, and projected impact of multimorbidity, both globally and in the UK, are described. This is followed by a description of key behavioural risk factors for multimorbidity that could serve as viable targets for preventative interventions. Here, the evidence on the clustering of four key risk behaviours, the potential for changes in behavioural patterns as we age, and their implications for epidemiological research are discussed. Finally, relevant risk behaviours are discussed in the context of multimorbidity and examined through the lens of social determinants of health.

2.2 Multimorbidity

2.2.1 What is multimorbidity?

In epidemiological research, multimorbidity is typically defined as the co-occurrence of two or more health conditions within an individual (Johnston et al., 2019). Although there is no universally agreed-upon definition of multimorbidity, it is different from — but often confused with — comorbidity (Gulliford and Green, 2024). Comorbidity involves multiple health conditions alongside an index condition, which is implicitly or explicitly treated as the main condition of concern, whereas multimorbidity encompasses the co-occurrence of multiple chronic medical conditions without specifying an index condition (Harrison et al., 2021). This distinction is pivotal in the clinical setting, where comorbid patients are evaluated in the context

of the index condition, while multimorbid patients' outcomes consider the interaction and burden of all co-existing chronic conditions (Harrison et al., 2021).

Moreover, the significance of identifying specific long-term conditions and their interactions in multimorbidity cannot be overstated. Rather than merely classifying patients as "multimorbid," it is essential to recognise how different combinations of conditions impact an individual's quality of life, responsiveness to medical interventions, and healthcare utilisation (Johnston et al., 2019). For example, a 40-year-old male with type 2 diabetes and depression is different to a 70-year-old female with osteoarthritis and early-stage Alzheimer's disease. Each combination of conditions carries distinct implications for the patient's quality of life and responsiveness to medical intervention. Thus, in the clinical context, the specific number is not as important as the concept of a patient with multiple conditions that can impact their long-term prognosis, healthcare utilisation and care trajectory (Gulliford and Green, 2024).

Similarly, multimorbidity may have a greater impact on overall health if it involves chronic conditions across different body systems that are likely to compete for treatment rather than closely related conditions that might have shared pathophysiology or shared approaches to management (Skou et al., 2022; Piette and Kerr, 2006). The construct of complex multimorbidity, defined as "the co-occurrence of three or more chronic conditions affecting three or more different body systems within one person without an index chronic condition", addresses these issues by focusing on chronic conditions affecting multiple body systems (Harrison et al., 2014, p. 1). This definition has the added advantage of identifying the number and types of specialised healthcare services required for a patient's care, thus identifying individuals with more complex needs (Harrison et al., 2014). Finally, complex multimorbidity

might also better reflect the biology of ageing as it involves a simultaneous breakdown or dysfunction of multiple, separate body systems (Singer et al., 2019a).

2.2.2 How common is it globally?

The prevalence of multimorbidity has risen in many regions of the world over the past 10 to 20 years, and it is anticipated to continue rising (Academy of Medical Sciences, 2018). Within the primary healthcare services of most high-income countries, multimorbidity is now the clinical norm (Violan et al., 2014). However, estimating its prevalence in populations is difficult not only because of variable levels of access to healthcare and rates of chronic condition diagnosis across the world but also because the prevalence of multimorbidity varies as per the definition of multimorbidity used, characteristics of the cohort under study (such as age, sex and socio-economic status) and country of study (Agborsangaya et al., 2012).

Nonetheless, global estimates of multimorbidity prevalence are high with a wide variation. A meta-analysis examining the global prevalence of multimorbidity in all age groups estimated it at 42.4% (Ho et al., 2022a). Another meta-analysis estimated the global prevalence of multimorbidity in community settings at 33.1% (Nguyen et al., 2019). Both studies reported large variability in estimates across included studies (Nguyen et al., 2019; Ho et al., 2022a). Similarly, a systematic review found the multimorbidity prevalence in primary care settings to range between 12.9% and 95.1% (Violan et al., 2014).

2.2.3 How it is measured and why it matters

The variability in estimating multimorbidity prevalence stems largely from the lack of international consensus on its definition. Prevalence estimates are strongly influenced by the

choice of operational definition of multimorbidity, which has two components: i) the list of diagnoses considered (i.e. the total number and types of conditions considered), and ii) the cut-off used to define the presence of the multimorbidity diagnosis (2+, 3+, 4+ diseases and so on). Regarding the former, there is currently no universally agreed-upon list of conditions (Ho et al., 2022b). And conditions span various categories, such as diseases, risk factors (e.g., obesity and hypertension), symptoms (e.g., chronic pain), and impairments (e.g., vision and hearing loss). Further, some researchers advocate for the inclusion of both acute and chronic conditions (Ho et al., 2022b).

While these factors contribute significantly to variations in multimorbidity prevalence, the number of conditions assessed seems to be the most critical issue in studying prevalence estimates. A review of 21 prevalence studies (both in primary care and among the general population) found much less variation in prevalence estimates when studies analysed 12 or more diseases compared to studies with a smaller number, and therefore, suggest using a list of at least 12 chronic diseases (Fortin et al., 2012). Although further research is needed to select which specific conditions to include, a list of the 12 most prevalent chronic diseases with a high impact or burden in a given population has been suggested as a good compromise. Similarly, with regards to the cut-off, a simple count of conditions with a cut-off at two or more values remains the most commonly used method for the measurement of multimorbidity (Fortin et al., 2012), yet the use of at least two operational definitions of multimorbidity to enhance comparison of findings across populations and developing universal guidelines and interventions has been suggested (Johnston et al., 2019).

2.3 Multimorbidity : A Focus on the United Kingdom

2.3.1 UK prevalence estimates and future projections

The prevalence of multimorbidity is poised to rise significantly in the United Kingdom, driven by factors such as an ageing population and declining mortality rates. Recent projections highlight the impending scale of this issue. Between 2015 and 2035, the percentage of adults aged 65 and over with multimorbidity is projected to surge from 54.0% to 67.8% by 2035 (Kingston, Comas-Herrera and Jagger, 2018). Moreover, numbers of the very old, aged 85 years and over, are set to double over the next 15 years (Nash, 2017). Multimorbidity is the norm in this age group, exacerbating the complexity of care, the risk of hospitalisations and re-admissions, and diminishing overall quality of life (Collerton et al., 2009). Similarly, a cross-sectional study from medical practices in Scotland revealed that by age 50 years, half of the population had at least one morbidity, and by age 65 years, most were multimorbid (Barnett et al., 2012).

Multimorbidity is not only manifesting earlier in the life course but also with a higher prevalence in recent cohorts (Head et al., 2020; Divo, Martinez and Mannino, 2014; Kingston, Comas-Herrera and Jagger, 2018). For instance, the Twenty-07 cohort study in Scotland reported a 60% increase in the prevalence of multimorbidity at age 60 in the cohort born in the 1950s compared with the cohort born in the 1930s (Katikireddi et al., 2017). These trends are expected to continue over the next two decades (Kingston et al., 2018). While age remains the primary driver of multimorbidity, more individuals under the age of 65 are affected by multimorbidity in absolute numbers compared to those aged 65 and above, largely due to the larger proportion of this age group within the population (Global Burden of Disease Study Collaborators, 2015; Barnett et al., 2012). Consequently, the contribution of younger cohorts

with multimorbidity as they age into the older population, along with growing numbers of the very old, could dramatically increase the health and social care burden in the coming years (Soley-Bori et al., 2021).

2.3.2 Impact and Cost

Multimorbidity is a costly condition that has now become the clinical norm in primary care, imposing a burden on patients, healthcare systems, and societies alike (Palladino et al., 2016; Picco et al., 2016). For patients, the functional challenges involve managing the complexity of multiple chronic conditions, physical limitations (i.e. pain or chronic fatigue), financial burdens, complexity of communication with healthcare providers, inadequate social support, logistical challenges in scheduling or coordination with healthcare providers, lifestyle changes, and the burden of multiple treatments (e.g., polypharmacy; Chen, Karimi and Rutten-van Mölken, 2020). Multimorbidity is strongly associated with higher mortality, poorer quality of life and functional status, and higher rates of health service use, including emergency hospital admission (Skou et al., 2022; Salisbury et al., 2011).

In the UK, the current single-disease oriented model of care delivery struggles to address the needs of patients with multimorbidity, who often experience care fragmentation, difficulty in managing their treatments, and poor health outcomes (Palladino et al., 2016). Better management of people with multimorbidity is therefore a key challenge for the current healthcare system.

Similarly, multimorbidity impacts society by straining resources and support structures, imposing significant challenges on healthcare systems. For instance, a recent meta-analysis of mostly high-income income countries reported that the annual costs of multimorbidity per

person ranged from I$800- I$150,000 (international dollars), depending on the country (increased country GDP per capita was associated with higher costs of multimorbidity), disease combination, cost ingredients, and other study characteristics (Tran et al., 2022). Another UK-specific review identified 17 studies that explored how multimorbidity affects healthcare costs and utilisation and suggested that multimorbidity translates to increased healthcare costs and utilisation (Soley-Bori et al., 2021). These included hospital costs, care transition costs, primary care use, dental care use, and emergency department use (Soley-Bori et al., 2021). However, the most sizeable effect of multimorbidity was seen on unplanned, potentially preventable hospitalisations, with up to 14.38 times increased odds of such hospitalisations for those with four or more conditions compared to those with none, an effect that was independent of age (Soley-Bori et al., 2021). In particular, the total primary care costs of multiple conditions were not strictly additive and instead depended on the specific disease combinations and age groups (Soley-Bori et al., 2021). Conversely, a systematic review that evaluated evidence from 16 European countries to investigate the relationship between multimorbidity and healthcare costs and utilisation showed that costs and utilisation (including physician visits, hospitalisations, and medication use) tend to increase with the number of conditions (Lehnert and König, 2012).

2.3.3 The case for multimorbidity prevention

The increasing prevalence of multimorbidity among both older and younger populations underscores an urgent need for proactive interventions and resource allocation to reduce the escalating burden of multimorbidity (Head et al., 2021a). Left unaddressed, the costs of treating multimorbidity within the current health and social care systems can become unsustainable and compromise the quality of patient care (Soley-Bori et al., 2021). Mitigating this burden will require effective strategies for prevention and early intervention. While the accumulation of

disparate diseases in multimorbidity makes prevention difficult if each disease is targeted separately, identifying and targeting shared risk factors for common diseases holds promise.

2.4 Behavioural risk factors

2.4.1 Behavioural risk factors as intervention targets

Extensive research has explored the role of lifestyle factors or health risk behaviours, such as smoking, alcohol consumption, poor nutrition, and physical inactivity (collectively termed 'SNAP' behaviours) in preventing individual chronic conditions (Bauer et al., 2014; Schmidt et al., 2016). SNAP behaviours account for nearly one-third of disability-adjusted life years from chronic conditions and have been suggested as key intervention targets by the World Health Organization (WHO) to lower the burden of chronic disease (Global Burden of Disease Study Collaborators, 2017). Notably, SNAP risk behaviours are the leading causes of major, non-communicable diseases worldwide and are also well-established risk factors for the "big four" chronic diseases contributing substantially to the multimorbidity burden—type-2 diabetes, coronary heart disease, chronic obstructive pulmonary disease, and specific cancers (Niebuur et al., 2023).

Given the well-established causal associations between SNAP risk factors and individual chronic diseases that comprise multimorbidity, it logically follows that focusing on SNAP risk behaviours as discrete targets for intervention presents a viable opportunity for multimorbidity prevention for a number of reasons (Nulu, 2017). First, targeting these behaviours is important since research identifies SNAP risk behaviours as standalone risk factors for multimorbidity itself (Álvarez-Gálvez et al., 2023; Skou et al., 2022). For instance, studies have observed associations between some unhealthy behaviours (smoking, alcohol consumption, physical

inactivity or diet quality) and prevalent cardiometabolic multimorbidity (Sakakibara, Obembe and Eng, 2019), prevalent multimorbidity (Fortin et al., 2014; Sakib et al., 2019), incident cancer and cardiometabolic disease multimorbidity (Freisling et al., 2020), as well as incident multimorbidity (Wikström et al., 2015). Equally, recent evidence on the longitudinal association of these health risk behaviours and the incidence of multimorbidity found a dose–response association between unhealthy lifestyle factors and multimorbidity, finding that a combination of two, three or four or more unhealthy lifestyle factors significantly increased the multimorbidity hazard, compared with none, from 42% to 116% (Dhalwani et al., 2017).

Secondly, focusing on these behaviours is not only likely to be helpful for primary prevention but also has the potential for improving outcomes in multimorbid patients (Wister et al., 2022). For instance, a study conducted in the UK found an inverse dose-response association between physical activity and mortality in individuals with multimorbidity as well as those without (Chudasama et al., 2019). Meeting the current physical activity guidelines, or as little as 10 min of brisk walking a day, was associated with a longer life expectancy in both sets of individuals (Chudasama et al., 2019). Similarly, healthier behaviours also play a crucial role in fostering multimorbidity resilience — defined as the ability to adapt to challenging life circumstances, including ageing-related changes, illness, or trauma (Seong et al., 2022). Multimorbidity resilience is associated with positive health outcomes despite the presence of chronic illnesses (Seong et al., 2022). A recent study found that among older adults with multimorbidity, healthier behaviours such as not smoking, good sleep quality, regular meals, and maintaining a healthy weight were associated with multimorbidity resilience (Wister et al., 2022).

Finally, targeting SNAP behaviours is important not only to improve health outcomes in older adults but also to prevent disease in the younger generations and those of working age. A recent

study has highlighted a concerning trend in England: a higher prevalence of both obesity and multimorbidity in younger cohorts compared to their counterparts a generation ago (Opazo Breton and Gray, 2023). Given that obesity has been identified as an independent risk factor for multimorbidity in those under 65 years of age, targeting health risk behaviours which contribute to obesity, such as SNAP behaviours, is likely to be beneficial (Delpino et al., 2023; Dhalwani et al., 2017). Equally, a recent systematic review of prevalence studies on multimorbidity reported a low prevalence before age 40 followed by a steep increase and a plateau after age 70 (Fortin et al., 2012). This suggests that prevention strategies can and should be used to intervene at all ages and that there is a considerable window of opportunity for primary prevention strategies to target a large population at risk.

2.4.2 Risk behaviours cluster together

To date, most intervention programmes have targeted single risk behaviours rather than considering a wider set of interlinked behaviours, particularly in older adults and young adults (aged 16-25 years) (Nigg and Long, 2012; Noble et al., 2015). However, health-risk behaviours rarely occur in isolation. Rather they co-occur or cluster together (Noble et al., 2015).

Health-risk behaviours are said to co-occur when they are prevalent in combinations in the population. For example, smoking has been found to co-occur with alcohol consumption (Nichter et al., 2010) and unhealthy diet (Meader et al., 2016). Similarly, physical inactivity often co-occurs with an unhealthy diet (Tassitano et al., 2014). In addition to evidence that health-risk behaviours often co-occur, epidemiological studies have also shown that such behaviours might co-occur more often than expected by chance, and instead cluster together in certain subgroups of the population (McAloney et al., 2013). These groupings are often referred to as ‘clusters’, and behavioural clustering demonstrates the tendency for certain combinations

of behaviour to group together into various lifestyles in certain individuals. In turn, identifying whether and how behaviours are closely related to each other is valuable because behaviours that share strong empirical relationships are more likely to have similar determinants (aetiologies; McAloney et al., 2013; Meader et al., 2016). For instance, the observation that 'risky' health behaviours are most common among more socio-economically disadvantaged individuals has been understood to mean that shared social determinants give rise to groupings of unhealthy (and healthy) behaviours (Marmot et al., 2012; Rose, 2001; Spring, Moller and Coons, 2012).

However, a key challenge in the research on multiple health-risk behaviours is the inconsistent terminology used to describe the interrelations between behaviours (McAloney et al., 2013). The terms clustering and co-occurrence are used interchangeably alongside similar terms such as bundling or concurrence (Spring, Moller and Coons, 2012). This creates difficulties in navigating the research because these terms are often substituted for one another despite some researchers arguing they mean different things. However, this thesis will rely on the distinction between co-occurrence and clustering provided by a recent scoping review of statistical approaches for analysing multiple health-related behaviours (McAloney et al., 2013). Co-occurrence refers solely to how prevalent a given combination of behaviours is, assuming patterns in those behaviours are independent of one another. Clustering of health-risk behaviours—in distinction to their co-occurrence — implies that the combinations of behaviours observed are not independent of each other and may, therefore, reflect an underlying causal mechanism. In other words, clustering concerns itself with underlying associations between concurrent behaviours, which co-occurrence does not. This distinction is reflected in the statistical analyses used to estimate each. Co-occurrence is estimated from "the prevalence of behavioural combinations and/or … summing behaviours into risk indexes"

(McAloney et al., 2013, p. 2). On the other hand, clustering can be recognised if the observed prevalence of a combination of behaviours is greater than the prevalence that would be expected if the behaviours were independent of each other (McAloney et al., 2013). Clustering can also be detected using advanced statistical methods capable of identifying latent (i.e., unobservable) clusters that are characterised by distinct underlying patterns in risk behaviours (McAloney et al., 2013).

Examining how health-risk behaviours cluster is important for three main reasons. First, the presence of multiple health-risk behaviours appears to create synergistic effects, including risks (i.e. an overall risk greater than the sum of the risks associated with each individual behaviour, indicating the behaviours act as effect modifiers for each other; Dhalwani et al., 2017; Marrero et al., 2005; Freedman et al., 2006; Mello et al., 2019). For instance, the Japan Collaborative Cohort study found that a composite measure of multiple health-risk behaviours predicted preventable death, over and above the predictive value of any single behaviour (Tamakoshi et al., 2009). Similarly, the consumption of both alcohol and tobacco (either smoked or smokeless) significantly increases the odds of oral cancer (Mello et al., 2019).

Secondly, since multiple health-risk behaviours can have synergistic effects on adverse health outcomes, simultaneously targeting such behaviours may offer the potential for increased health benefits and reduced healthcare costs (Prochaska et al., 2010). However, a recent systematic scoping review that examined the characteristics of interventions targeting multiple health-risk behaviours described a divergence between the evidence on the relationship between multiple risk behaviours and the combinations of behaviours chosen as targets of such interventions (King et al., 2015). For example, the review found consistent evidence indicating that alcohol misuse and smoking clustered together in adults across large-scale surveys

conducted in Scotland, England, and the Netherlands (Poortinga, 2007; Shankar, McMunn and Steptoe, 2010; King et al., 2015). However, not a single study among the 220 interventions mapped in the scoping review targeted alcohol misuse and smoking together (King et al., 2015). By highlighting how the behavioural combinations targeted by multiple risk behaviour interventions may not reflect consistent research findings on the clustering of multiple health-risk behaviours, the review by King et al. (2015) highlights a notable discrepancy between research and practice. For more convergence between research and practice, it can be argued that more compelling evidence on how health-risk behaviours cluster in specific populations and the health impacts of such clusters may be required.

Finally, in the context of intervention design, it is important to understand how health-risk behaviours cluster since findings have provided support for the presence of transfer effects, wherein health-promoting and health-risk behaviours correlate within behaviour clusters, but not between these groupings (Lippke, Nigg and Maddock, 2012). In other words, research into clustering among health behaviours "suggests substitute and complementary relationships, such that changing one behaviour may crowd out or make room for interconnected behaviours or, alternatively, bring along…same direction changes in linked actions" (Spring, Moller and Coons, 2012, p. 6). An individual-level example of behaviour substitution is the finding that when an individual reduces cigarette smoking, the reward value and consumption of 'treat foods' increases, culminating in weight gain (Audrain-McGovern and Benowitz, 2011). It is presumably for this reason that attempting to simultaneously rein back on two rewarding substances (e.g., cigarettes and treat foods) is less effective than making each behaviour change sequentially with a brief time lag (Farley et al., 2012). There is also population-level evidence for changes in one behaviour leading to changes in other behaviours, or perhaps, trends in two behaviours suggesting a common underlying cause. For instance, complementary behaviour

change has been observed when the early onset of smoking tends to be followed by the onset of other health-risk behaviours like illicit drug use — a phenomenon known as the gateway hypothesis (Miller and Hurd, 2017). Thus, a better understanding of how health-risk behaviours cluster may allow intervention designers to pre-empt the effect of interventions on related behaviours that may not be directly targeted by such interventions.

In addition to the broad benefits of understanding how health-risk behaviours cluster in the general population, it is important to study how such behaviours cluster in older adults. This is because, despite evidence indicating the potential benefits of interventions targeting multiple health-risk behaviours, such interventions rarely focus on older adults. This is evident in a scoping review on interventions targeting multiple health-risk behaviours, which found that out of the 100 interventions that focused on subgroups of general adult populations, only six targeted older adults. Thus, more evidence to understand the clustering of health-risk behaviours would be useful, especially in a population that has not been given much attention in this context (i.e. older adults; King et al., 2015).

In sum, measuring and monitoring how health-risk behaviours cluster could be a valuable tool for summarising the patterns of participation in key health-risk behaviours seen across a population and in helping to identify key intervention points (e.g., over-represented behavioural patterns within specific subgroups). Understanding how health-risk behaviours cluster can help make interventions more efficient by directing them at population subgroups exhibiting the highest aggregation of risk factors.

More importantly, in light of the UK's ageing population, rising prevalence of chronic diseases and multimorbidity, and escalating healthcare costs, understanding how individuals cluster in

terms of their engagement in health-risk behaviours can inform the design of age-appropriate health interventions (United Nations & Social Affairs, 2015).

2.4.3 Current evidence on the clustering of health-risk behaviours

Research on the clustering of health-risk behaviours encompasses a recent yet rapidly growing body of work. Two recent systematic reviews have examined the clustering of multiple health-risk behaviours in the general adult population (Meader et al., 2016; Noble et al., 2015). While Noble et al. (2015) took an international perspective, Meader et al. (2016) focused on studies from the UK. A third review examining the clustering of multiple health-risk behaviours focused on patterns of clustering in children between the ages of 11–16 (Whitaker et al., 2021).

In the general adult population, both 'healthy' and 'unhealthy' clusters were found to be common, with larger numbers of individuals than would be expected by chance exhibiting all, or none, of a range of health risk behaviours (Meader et al., 2016; Noble et al., 2015). Notably, the most frequently reported clusters in the review by Noble et al. (2015) were characterised by the absence of any SNAP risk behaviour (observed in 81% of the included studies). This was followed by the clustering of smoking and alcohol consumption only (observed in 56% of the studies), the presence of all four SNAP risk behaviours (50% of studies), and low physical activity with poor diet (44% of studies). Similarly, Meader et al. (2016) noted strong evidence for the clustering between smoking and alcohol misuse, and smoking and unhealthy diet.

Among young adults, smoking, alcohol misuse, and illicit drug use were consistently found to cluster, while a minority of young people in many of the included studies exhibited none of the included health-risk behaviours (Meader et al., 2016). This polarisation between engaging in multiple unhealthy behaviours or no unhealthy behaviours was also apparent in young children,

as most 11–16 -year-olds fell into one of the 'healthy' clusters, which, on average, accounted for 51% (Substance Abstainers) or 32% (Overall Healthy) of the primary study populations (Whitaker et al., 2021). However, small minorities were reported to engage in all health risk behaviours (Whitaker et al., 2021).

Notably, all three of the reviews also found a large number of clusters that were more difficult to characterise. For example, Whitaker et al. (2021) found clusters in several studies that involved unexpected patterns of engagement in health-risk behaviours (e.g., engagement in both high levels of physical activity and unsafe sex). On the other hand, Meader et al. (2016) only described the most common behavioural clusters and did not comment on clusters that were found to be ambivalent or contradictory across the included studies.

2.4.4 Key issues in the literature on risk behaviour clustering

Importantly, each of the aforementioned systematic reviews highlighted two issues that create problems in comparing data across studies examining the clustering of health-risk behaviours: i) methodological diversity in the types of clustering analyses used and ii) differences in measures and definitions of health-risk behaviours. The first issue regarding the diversity in the types of clustering analyses used emphasises a lack of consensus in how clustering itself is defined (McAloney et al., 2013). For instance, some methods identify clustering on the basis of divergences between observed and expected prevalence. Other methods (e.g., latent class analysis) identify common patterns in participant responses to the behaviour items and thus allow behaviours to be categorised into underlying typologies, termed ‘clusters’ or ‘classes’ (McAloney et al., 2013).

Nonetheless, the issue of methodological diversity in the types of clustering analyses used can be hard to resolve since different statistical approaches to clustering have advantages and disadvantages that need to be weighed against the data characteristics and research questions. This can make it hard to declare one methodology unambiguously superior to another. For example, two common statistical methods for analysing behavioural clusters are factor analysis (Yong and Pearce, 2013) and latent class analysis (Scotto Rosato and Baer, 2012). Since the former typically requires continuous data and the latter uses categorical data (although both have extensions that can handle wider ranges of data types), the type of data that is available in the first place plays an important role in choosing the type of analysis. This, in turn, determines the type and content of the results obtained from the analysis.

The second issue, however, is more amenable to change. This issue refers to the diversity in the measures and definitions of health-risk behaviours across studies that aim to examine their clustering (Noble et al., 2015). Such diversity limits opportunities for comparing study findings and drawing inferences about which health-risk behaviours can be expected to cluster and for which subgroups. For instance, measures of alcohol included in studies examining clustering patterns of 'SNAP' behaviours in the general adult population include a week-long diary (Bondy and Rehm, 1998), quantity-frequency measures (De Vries et al., 2008), a diary for the last two 48-hr drinking periods (Burke et al., 2007), number of glasses of alcohol per week (Lucini et al., 2011), number of days of drinking in the last 30 days (Quintiliani et al., 2010), binge drinking frequency (Rebholz et al., 2012), and the Alcohol Use Disorders Identification Test (AUDIT) scores (Tobias et al., 2007). Similar variation has also been observed in the context of diet and physical inactivity (Noble et al., 2015). Lesser variation has been observed in how clustering studies define smoking, with a majority assessing current smokers versus non-smokers. However, it is not uncommon for studies examining clusters of health-risk

behaviours to include a measure of cigarette consumption (De Bourdeaudhuij and Van Oost, 1999), define smoking in pack-years (McLernon et al., 2012), or define smokers strictly as those who smoke daily (Lv et al., 2011), or those that smoke more than 5 cigarettes per day (Fu, 2004).

Inconsistency in definitions and measures of health-risk behaviours can result in estimates of behaviour that are not directly comparable and make it difficult to draw conclusions from available evidence to inform intervention design or other policies (McAloney et al., 2013). Therefore, systematic reviews examining how health-risk behaviours cluster highlight the need for expert consensus on the definition of SNAP health-risk behaviours as it would allow researchers to conduct more detailed comparisons and make reliable inferences (McAloney et al., 2013; Meader et al., 2016; Noble et al., 2015; Whitaker et al., 2021)

2.4.5 Clustering of health-risk behaviours over time

Most of what is currently known about the clustering of health-risk behaviours comes from cross-sectional studies, where the data only provides a snapshot of the behavioural patterns (Noble et al., 2015). However, health behaviours are fairly dynamic across time. Several studies have identified distinct developmental trajectories for single health-risk behaviours like physical exercise (Lounassalo et al., 2019), and alcohol consumption (Tucker, Orlando and Ellickson, 2003), indicating that health behaviours change across the lifetime. For instance, Lounassalo et al. (2019) found several trajectories that indicated a decline in physical activity over time in the youngest group, whereas trajectories of steadily rising physical activity were found among the middle and the oldest age groups. However, the proportion of persistently physically inactive individuals was reported to increase with age, even as the proportion remained relatively high across all ages.

There is also evidence to suggest that the type and number of clusters vary across age (van Nieuwenhuijzen et al., 2009), likely reflecting varying physical and social contexts or different levels of engagement in health-risk behaviours across ages. Among general adult populations, for instance, smoking and alcohol misuse was the most commonly identified cluster across the systematic reviews discussed above (Meader et al., 2016; Noble et al., 2015). In contrast, for young adults, the most consistent evidence was found for the clustering of sexual risk behaviour with illicit drug use, smoking and alcohol misuse (Meader et al., 2016). Importantly, systematic reviews examining the clustering of health-risk behaviours in adults have reported difficulties in determining which behaviours were more likely to cluster among older adults due to a lack of studies (Meader et al., 2016; Noble et al., 2015). In other words, there is limited understanding of which health-risk behaviours cluster together in older adults.

Given the points above, it is reasonable to assume that health behaviour clusters could also evolve as people age. However, to our knowledge, few studies have attempted to examine how multiple health-risk behaviours cluster over time (Buck and Frosini, 2012; Daw, Margolis and Wright, 2017; Goodhind, Gilchrist and Memon, 2014; Hsu et al., 2013; Feng et al., 2022). Although two of these five studies (Buck and Frosini, 2012; Goodhind, Gilchrist and Memon, 2014) claim to examine the clustering of health-risk behaviours over time, the methods they have used at best describe changes in the co-occurrence of behaviours over time. Moreover, the two studies looked at two separate years of a cross-sectional survey and, as such, did not follow the same individuals across time; rather, they observed two different populations sampled using the same methods. In doing so, the studies examined period effects (i.e., how the population as a whole is changing over calendar time) but giving us little insight into age effects (i.e. how changes occur as individuals age; Holford, 1992).

Therefore, to date, there appear to be just three studies that have examined how multiple health-risk behaviours cluster as individuals age. Two studies examined longitudinal patterns in multiple risky behaviours among older adults (aged ≥ 50 years) in China (Feng et al., 2022) and Taiwan (Hsu et al., 2013), respectively. The study in China examined three health risk behaviours, namely smoking, physical activity and social participation, examined over three waves of data collection (i.e. across four years) of the China Health and Retirement Longitudinal Study (Feng et al., 2022). The study found seven clusters with distinct behaviour patterns which were mostly stable over time, which are not directly comparable to the clusters found in previous reviews on account of differences in behaviours examined, however the study did find a consistently healthy cluster and unhealthy cluster (Meader et al., 2016; Noble et al., 2015). Similarly, the study in Taiwan examined four health risk behaviours, namely smoking, alcohol consumption, physical activity, and health check-ups, examined over three waves of data collection (i.e. twelve years) in the Taiwan Longitudinal Survey on Ageing (Feng et al., 2022). The study found that trajectories of health-related behaviours differed between males and females, but on the whole the patterns of behaviours in all clusters were mostly stable over time with one exception: some clusters exhibited a reduction in alcohol consumption and smoking over time. Five groups of older males were identified: smoking, inactive, healthy lifestyle, smoking and drinking, and quitting. In contrast, three groups of older females were identified: smoking and drinking, inactive, and healthy lifestyle. Notably, none of the aforementioned studies examined the clustering of SNAP behaviours.

While the aforementioned studies focused on older adults, the third study focused on a younger population (Daw, Margolis and Wright, 2017). The study by Daw, Margolis and Wright (2017) used four waves of longitudinal data to chart how four health-risk behaviours — cigarette smoking, alcohol consumption, obesity and sedentary behaviour — cluster over the transition

from adolescence (ages 13-18) to adulthood (ages 24-32). The study found several groups with consistently healthy or unhealthy behaviours. For example, the consistently healthy group (19.2% of the sample) and the least healthy group (8.5% of the sample) comprised participants who had fairly stable trajectories of behaviours across the transition from adolescence to adulthood. The healthiest group was found to be consistently healthy relative to other classes in all four behaviours. However, the least healthy group was not the unhealthiest across the four health behaviours; instead, it had the highest eventual rate of obesity but not the highest rates of either binge drinking, smoking, or sedentary behaviour. Most of the other groups (making up more than two-thirds of the sample) comprised participants who experienced some change over the transition to adulthood, with most of such change indicating a trend towards less healthy profiles. Thus, in studying the longitudinal clustering of health behaviours, the research by Daw, Margolis and Wright (2017) paved the path for investigating clustering of multiple health-risk behaviours across the life course.

However, Daw, Margolis and Wright (2017) highlighted three issues that need to be examined to better understand how health behaviours cluster across time, especially if behavioural clusters are to be examined in samples with a wider age range. First, although Daw, Margolis and Wright (2017) examined behaviours, they included obesity as an indicator of dietary behaviour. A more appropriate and informative variable to uncover underlying dietary patterns would be nutrition, which remains untested in this context. Secondly, to examine drinking behaviour, the study only assessed whether or not an individual engages in any episodes of binge drinking. Although binge drinking — high alcohol intake in a single drinking session— is the most frequent alcohol consumption pattern in youth, it is unlikely to accurately capture drinking patterns among older adults, which are more heterogeneous with regards to frequency and volume of intake Lovatt et al. (2015).

Finally, Daw, Margolis and Wright (2017) modelled each health risk behaviour as a dichotomous variable, thus only testing whether or not an individual was engaging in a risky behaviour at each point in time and not the intensity with which they did so. Since the study examined patterns of behaviour across the transition from adolescence to adulthood – a period recognised for marked and abrupt change in life circumstances and health behaviours – this strategy was likely appropriate for capturing patterns of behaviour change (Lenz, 2001). Indeed, prior research on behaviours like alcohol intake and physical activity has consistently indicated that changes in health behaviours are far fewer and more subtle in later stages of life (i.e., mid- to older adulthood) as compared to the changes observed across earlier life stages (e.g. from adolescence to young adulthood). One study, for instance, found that physical activity in middle age (ages 42-52 years) was associated with physical activity after 13 years of follow-up, suggesting that health behaviour patterns in mid-adulthood partly track into old age (Gabriel et al., 2017). Equally, research examining alcohol intake over a 24-year period in a cohort of older adults observed that drinking patterns remained relatively stable from middle to older age (McEvoy et al., 2013). Given the fairly stable population-level health behaviour trajectories of older age groups, it follows that examining the intensity and frequency with which a behaviour is performed is likely to better capture the underlying within-individual differences over time than binary responses that record engagement in or abstinence from a given behaviour (Rhodes et al., 2017).

More generally, there are a couple of reasons why clustering is likely to differ across the life course. First, the impact of life events likely to cause a change in lifestyles tends to be markedly dissimilar between young, middle-aged, and older adults. For instance, a general trend toward increased drinking and smoking upon entering college is a common phenomenon among

adolescents and young adults (Hensel, Todd and Engs, 2014). However, the directional impact of life events like bereavement, retirement, and children leaving home on health-risk behaviours may be more variable and subtle (Zantinge et al., 2014). Secondly, since clustering of health-risk behaviours is known to have synergistic effects, it is likely that such synergies will manifest particularly strongly in older age, where cumulative health effects of risk behaviours are particularly pronounced. The increased risk of harm in older age groups may also mean an increased likelihood that people moderate their behaviour as a preventative measure (Wiscott, Kopera-Frye and Begovic, 2002).

2.4.6 Clustering of health-risk behaviours and disease outcomes

Taking a life course approach to examine health risk behaviours and whether and how they cluster is valuable as it acknowledges that measuring risk factors only once may not reflect their full impact on morbidity and mortality through the life course (Burke et al., 2007). Presently, most of what is known about the clustering of health-risk behaviours and their association with chronic diseases comes from cross-sectional studies (e.g., Griffin et al., 2014). Cross-sectional studies provide little insight into the long-term health effects of different behavioural cluster profiles. Thus, it is not possible to ascertain whether a causal relationship exists between health behaviour patterns and health outcomes (Buck and Frosini, 2012). Longitudinal studies are needed to determine how different cluster patterns track over time and their influence on the development of specific chronic conditions.

Equally, studies examining risk behaviour clusters and their association with multimorbidity are lacking. Most studies on the association between multimorbidity and health-risk behaviours either investigate individual risk behaviours, interactions between two behaviours, or use a cumulative risk index by summing up the number of risky behaviours an individual engages in

(Fortin et al., 2014; Katikireddi et al., 2017; Martens et al., 2015). This offers a limited approach. Investigating single health behaviours is insufficient to provide insights into the distinct ways that multiple health behaviours combine into meaningful lifestyle patterns to influence disease outcomes. Similarly, a cumulative risk index does not tell us which combinations of behaviours drive the risk. Although such measures are useful when it comes to estimating population-attributable risk of disease, they are unlikely to provide key information for designing preventive interventions (i.e. whom and which behaviours to target; Shang et al., 2020). Other studies have tackled this issue by examining the health risks associated with combinations of behavioural factors in dyads, triads, and tetrads (Dhalwani et al., 2017). But they tend to overlook the rarer but riskier patterns of multiple risk behaviours because of sparse data (Dhalwani et al., 2017).

To date, no study has examined the longitudinal association between clusters of health behaviours over time and multimorbidity.

2.5 Social Determinants of Health (SDOH) Framework

2.5.1 Social gradient in risk behaviours and multimorbidity

Risky health behaviours that contribute to chronic diseases and multimorbidity, such as SNAP behaviours, tend to be socially patterned. They cluster within certain population groups (Noble et al., 2015). For instance, risky behaviours are more prevalent among men, economically inactive individuals, and people with low socio-economic status and educational attainment (Poortinga, 2007; Noble et al., 2015).

The presence of a social gradient in risk behaviours is also mirrored in multimorbidity prevalence and patterns. In addition to age, a key predictor of multimorbidity, demographic factors like sex are also linked to multimorbidity prevalence (Álvarez-Gálvez et al., 2023). A systematic review (Álvarez-Gálvez et al., 2023) of 39 studies including > 70 million patients found the prevalence of multimorbidity to be slightly higher in women than in men, which is consistent with previous findings (Skou et al., 2022). Notably, the review found distinct patterns of multimorbidity in men and women. For example, cardiovascular patterns were more prevalent in men, while musculoskeletal and mental patterns were common in women (Álvarez-Gálvez et al., 2023).

Evidence also suggests a strong socio-economic gradient in multimorbidity prevalence (Álvarez-Gálvez et al., 2023). A review examining household and area-level determinants of multimorbidity in high-income countries found that the odds of multimorbidity were up to 4.4 times higher for those within the lowest level of household income (versus the highest), and prevalence was 1.4 times higher in the most versus the least deprived areas (Ingram et al., 2021). Similarly, a study using a large representative study of adults in England and Wales found that multimorbidity prevalence was higher in the most deprived decile of deprivation (22%) compared to the least deprived decile (16% ; Fumagalli, Knies and Buck, 2017). Besides influencing multimorbidity prevalence, socio-economic factors also influence its onset. Multimorbidity manifests earlier in the life course for people living in more deprived areas compared to their affluent counterparts (Violan et al., 2014; Barnett et al., 2012).

More concerningly, a recent microsimulation study based on English primary care data projected an 84% surge in multimorbidity cases between 2019 and 2049, with the most

significant increase expected in the most deprived quintile of the population (Head et al., 2024). The disparity in multimorbidity burden is expected to widen over the coming decades, particularly among the working-age population (Head et al., 2024).

This is especially important considering a systematic review of interventions in primary care and community settings, which found minimal consideration of the impact of socio-economic deprivation in the included studies, and no study considered the possibility of a differential effect of interventions in different socio-economic groups (Smith et al., 2012).

Furthermore, it is crucial to acknowledge that multimorbidity is associated with social and economic determinants (Skivington et al., 2015) that can add to (or interact with) behavioural determinants.

2.5.2 Incorporating the SDOH framework

The Social Determinants of Health (SDoH) framework describes how broader societal structures – from economic policies to social norms – shape and segment populations hierarchically based on gender, race, education, occupation, and income (Fuchs, 2017). Social determinants of health refer to the conditions in which people are born, grow, live, work, and age, and the wider set of forces and systems shaping the conditions of daily life (Braveman and Gottlieb, 2014). These determinants can take the form of various social, cultural, economic, and political conditions that influence the health of individuals and populations and are crucial in understanding health inequities and disparities (Fuchs, 2017).

Given the social gradient observed in both risk behaviour and multimorbidity, there remains a need to better establish whether there are risk factors or behaviours which can be used as targets for preventive interventions and to confirm whether lifestyle management strategies need to be more optimally directed to certain sub-populations. Given that socio-demographic determinants can not only shape individual health behaviours but also predict health outcomes through complicated multifactorial pathways, I will adjust for them when examining the relationship between health behaviours and outcomes as well as assess how socio-demographic characteristics are independently associated with identified clusters (Northwood et al., 2018).

In this thesis, I have used the SDoH Framework to evaluate health inequalities by incorporating individual-level socio-demographic characteristics. As this thesis was primarily focused on clusters of multiple risk behaviours over time and disease, I used a simplified interpretation of SDoH to maintain the study's focus and to prevent the modelling from getting excessively complicated, thus hard to interpret. Although I recognise the significance of social determinants like area-level deprivation (such as inadequate and limited housing), environmental factors (including cultural norms and community support systems), and structural forces (such as resource access, discrimination, historical inequities, and systemic biases within institutions) in influencing health outcomes, the present work was unable to include them. As a result, the focus remained primarily on individual-level variables to examine social determinants of health inequalities.

2.6 Current gaps, aim, and research questions

This chapter has identified an important evidence gap related to our understanding of health risk behaviours that have implications for public health. The SNAP behaviours are leading risk factors for multimorbidity and tend to cluster (i.e. occur in specific combinations within distinct

subpopulations). However, little is known about how these clusters change with age in adults and older adults. Moreover, questions around risk behaviour clusters and disease have tended to rely on cross-sectional data. To our knowledge, no study has examined the longitudinal association between clusters of health behaviours over time and health outcomes like multimorbidity, i.e., the co-occurrence of two or more diseases.

More specifically, it remains unclear how SNAP risk behaviours cluster and change with age , what are the socio-demographic profiles of different clusters and how the clusters may be associated with multimorbidity. This information could provide us valuable information on any high-risk subgroups and better allow us to examine whether and how risk behaviour profiles link to disease.

2.6.1 Thesis aim

This thesis aims to advance our understanding of how SNAP risk behaviours cluster and change with age in the UK population. It also aims to investigate the socio-demographic profiles of the identified clusters and assess whether and how the clusters are prospectively associated with multimorbidity and a range of chronic disease outcomes.

2.6.1 Objectives

1)

a) Explore how the SNAP behaviours cluster over time in older adults in the UK.
b) Investigate how membership in different behavioural clusters varies by socio-demographic characteristics.

c) Examine which, if any, behavioural clusters are prospectively associated with multimorbidity over time.

2) Replicate the analysis in Objective 1 using a longitudinal panel of adults (aged 16 years and older) in the UK.
3) Investigate the possibility of socio-demographic characteristics moderating the relationship between risk behaviour clusters and disease outcomes in older adults.

Chapter 3: Methodological approach

In this section, I outline the considerations that drove the selection of analytical methods before explaining why Repeated Measures Latent Class Analysis (RMLCA) was chosen.

The selection of analytical methods in the subsequent chapters was driven by four main considerations.

First, previous studies examining behavioural clusters primarily utilised latent analytic and factor analytic approaches, which help reduce data dimensions to extract patterns from otherwise complex data (McAloney et al., 2013). Given the focus of this thesis on risk behaviour clustering, it was valuable to explore methods aligning with these established approaches.

Second, the nature of the data played a pivotal role in method selection. Upon cleaning and processing the chosen longitudinal datasets, it became evident that commonly used measures of SNAP risk behaviours, particularly for smoking and physical activity, were categorical. In light of this observation, it was important to select an analytical method capable of accommodating categorical data in longitudinal studies.

Third, I anticipated substantial missing data due to the longitudinal nature of the risk behaviour data. For this reason, it was crucial to select a method that could handle this issue.

Finally, the present research aimed to identify and link health behaviour clusters to relevant socio-demographic characteristics and disease outcomes, necessitating a methodological

approach focused on individuals rather than variables. A person-centred approach, which emphasises individual patterns and subgroup dynamics within a population, was preferred over a variable-centred approach. The latter primarily examines the relationships among variables across the entire sample or population, which is less effective for identifying distinct subgroups or exploring variations within individuals over time.

In light of these considerations, I chose Repeated Measures Latent Class Analysis (RMLCA) — an extension of latent class analysis — because it adeptly handles longitudinal, categorical data. Further, it is a person-centred approach that would allow distinct subgroups with common risk behaviour patterns to be identified. In other words, the emphasis in RMLCA is not on the variable (not even time) but on the person (i.e. what patterns emerge across persons over time).

A key advantage of using RMLCA, particularly using the MPlus software, lies in its Full Information Maximum Likelihood (FIML) capability. FIML addresses missing data comprehensively, by ensuring that parameter estimates are derived from all available information (He and Fan, 2018; Asparouhov and Muthén, 2014). Rather than disregarding or imputing missing data with specific values, it uses all available information while accounting for the uncertainty introduced by missing data. Such robustness is crucial given the expected volume of missing data on health risk behaviours in longitudinal datasets (Peeters et al., 2015). Furthermore, RMLCA's probabilistic model-based framework produces statistical estimates of model parameters, enabling robust hypothesis testing and providing other objective indices for comparing models (Dias, 2006).

Other methods considered but not chosen are briefly discussed below.

Growth curve modelling, while a person-centred approach suitable for longitudinal data, was ruled out because it cannot effectively handle categorical data and requires fitting specific functional forms to data trends. This approach is less flexible compared to RMLCA, which does not constrain the shape of the data trend — the trend over time can take any shape that fits the data, no matter how complex (He and Fan, 2018).

Latent transition analysis (LTA) was also considered (Hultgren et al., 2019). LTA enables the analysis of changes in class membership across different time periods. To date, most studies have studied the transition over time in one or (at most two) related behaviours. In the present case, four different health behaviours over six time points needed to be analysed since the impact of risk behaviours over one or two time points would be inadequate to measure the impact on disease outcomes. Conducting an LTA on such data, which would require demonstrating transition matrices across six time points, was unfeasible due to excessive complexity (Lanza, Flaherty and Collins, 2003). Moreover, in principle, latent transition analysis is more applicable to stage-sequential processes (e.g. examining smoking cessation patterns based on the transtheoretical model of change) where the transition between different stages can be tracked across time (Guo et al., 2009). However, given that the current data was unlikely to represent distinctive stages that could be estimated from prior theory or interpreted from resulting analysis, latent transition analysis was unlikely to be a good analytic strategy.

For these reasons, RMLCA was selected as the most suitable method due to its capacity for identifying and grouping individuals with similar patterns of health-risk behaviours over time into clusters (Dias, 2006).

Chapter 4: Longitudinal clustering of health behaviours and their association with multimorbidity in older adults in England: A latent class analysis

Overview: The first study of this thesis has been published in Plos one under Open Access (Suhag, Webb and Holmes, 2024). It is presented in this thesis in the format submitted.

Longitudinal clustering of health behaviours and their association with multimorbidity in older adults in England: A latent class analysis.

Authors: Alisha Suhag[1*], Thomas L. Webb[2], John Holmes[3]

[1] Healthy Lifespan Institute, School of Health and Related Research, University of Sheffield, UK

[2] Department of Psychology, University of Sheffield, UK

[3] School of Health and Related Research, University of Sheffield, UK

***** Corresponding Author:

E-mail: asuhag1@sheffield.ac.uk

Abstract

Background: Health-risk behaviours such as smoking, unhealthy nutrition, alcohol consumption, and physical inactivity (termed SNAP behaviours) are leading risk factors for multimorbidity and tend to cluster (i.e. occur in specific combinations within distinct subpopulations). However, little is known about how these clusters change with age in older adults, and whether and how cluster membership is associated with multimorbidity.

Methods: Repeated measures latent class analysis using data from Waves 4-8 of the English Longitudinal Study of Ageing (ELSA; n=4759) identified clusters of respondents with common patterns of SNAP behaviours over time. Disease status (from Wave 9) was used to assess disorders of eight body systems, multimorbidity, and complex multimorbidity. Multinomial and binomial logistic regressions were used to examine how clusters were associated with socio-demographic characteristics and disease status.

Findings: Seven clusters were identified: *Low-risk* (13.4%), *Low-risk yet inactive (*16.8%), *Low-risk yet heavy drinkers* (11.4%), *Abstainer yet inactive* (20.0%), *Poor diet and inactive* (12.9%), *Inactive, heavy drinkers* (14.5%), and *High-risk smokers* (10.9%). There was little evidence that these clusters changed with age. People in the clusters characterised by physical inactivity (in combination with other risky behaviours) had lower levels of education and wealth. People in the heavy drinking clusters were predominantly male. Compared to other clusters, people in the *Low-risk* and *Low-risk yet heavy drinkers* had a lower prevalence of all health conditions studied. In contrast, the *Abstainer but inactive* cluster comprised mostly women and had the highest prevalence of multimorbidity, complex multimorbidity, and endocrine disorders. *High-risk smokers* were most likely to have respiratory disorders.

Conclusions: Health-risk behaviours tend to be stable as people age and so ought to be addressed early. We identified seven clusters of older adults with distinct patterns of behaviour, socio-demographic characteristics and multimorbidity prevalence. Intervention developers could use this information to identify high-risk subpopulations and tailor interventions to their behaviour patterns and socio-demographic profiles.

Introduction

A growing number of older adults are living with multimorbidity, defined as having two or more chronic diseases (Kingston et al., 2018). In England, for example, 67.8% of people aged 65 and older are expected to have multimorbidity by 2035 (Kingston et al., 2018). Multimorbidity is more common in recent cohorts and seems to be emerging earlier in the lifecourse, which places a significant strain on healthcare systems (Canizares et al., 2018; Picco et al., 2016). However, chronic diseases (e.g., type-2 diabetes, coronary heart disease, chronic obstructive pulmonary disease) and some cancers that form a large proportion of the multimorbidity burden, have well-established modifiable risk factors, suggesting that there are opportunities for prevention (Barnett et al., 2012). Among such modifiable risk factors – health risk behaviours such as smoking, poor nutrition, alcohol consumption, and physical inactivity (collectively termed 'SNAP' behaviours) account for nearly one-third of disability-adjusted life years from chronic conditions (Collaborators, 2017). Most preventative interventions for older adults focus on single behaviours (Nigg and Long, 2012). However, research shows behavioural risk factors typically cluster in specific combinations within distinct populations. This suggests that interventions targeting these clusters may be more appropriate (Noble et al., 2015). However, designing such interventions is challenging because individuals' health behaviours not only cluster but may also change over time (Botoseneanu and Liang, 2012).

Additionally, engaging in multiple health-risk behaviours can have a negative impact on health that is greater than the sum of their individual effects (Tamakoshi et al., 2009). While epidemiological studies have attempted to understand the combined impact of multiple behaviours on health, they often rely on simple indices that count the number of co-occurring behaviours without considering which behaviours, or combinations of behaviours, are driving the risk (McAloney et al., 2013; Katikireddi et al., 2017). Other studies have tackled this issue by examining the health risks associated with combinations of behavioural factors in dyads,

triads, and tetrads (Dhalwani et al., 2017). However, they tend to overlook less common combinations involving multiple risk behaviours because of sparse data (Dhalwani et al., 2017). Clustering techniques, such as repeated measures latent class analysis (RMLCA), can better address these issues by grouping individuals with similar patterns of health-risk behaviours over time into clusters (McAloney et al., 2013).

Identifying clusters of health-risk behaviours – and examining whether and how behaviours within these clusters change as people age – can inform interventions by: a) identifying high-risk populations (e.g. subgroups exhibiting combinations of behaviours that entail the greatest risk), b) informing the selection of target behaviours (e.g. those with the greatest health impact or highest reach), and c) detecting potential spillover effects (i.e. where targeting one health behaviour leads to compensatory changes in other behaviours)(Filozof, Fernandez Pinilla and Fernández-Cruz, 2004). For instance, when some individuals reduce cigarette smoking, the reward value and consumption of 'treat foods' increases, resulting in weight gain (Filozof, Fernandez Pinilla and Fernández-Cruz, 2004).

Previous studies that have used clustering techniques to analyse multiple behaviours in older adults have typically been cross-sectional, and so cannot test whether behaviour clusters change over time and if these changes affect long-term health (Schneider et al., 2009; Griffin et al., 2014; Liao et al., 2019). While two studies have used longitudinal clustering to study a subset of SNAP behaviours in older adults (Hsu et al., 2013; Feng et al., 2022), none have examined the relationship between these clusters and multimorbidity. In addition, the aforementioned studies limit themselves to a basic definition of multimorbidity (i.e. a simple count of the number of diseases), which overlooks differences between diseases within one system and those spanning multiple systems (Wister et al., 2015). This is crucial, as

multimorbidity may have a larger impact on overall health if it arises out of chronic conditions in different body systems that are likely to compete for treatment, rather than closely related comorbidities that might have shared pathophysiology or shared approaches to management (Skou et al., 2022). The construct of complex multimorbidity, defined as "the co-occurrence of three or more chronic conditions affecting three or more different body systems within one person without an index chronic condition", addresses these issues by focusing on chronic conditions affecting multiple body systems (Harrison et al., 2014, p1). This definition also has the advantage of identifying the number and types of specialised health services involved in a patient's care, thus identifying individuals with more complex needs (Harrison et al., 2014). Complex multimorbidity might also better reflect the biology of ageing as it involves a simultaneous breakdown or dysfunction of multiple, separate body systems, making it a more reflective measure to study in older adults (Singer et al., 2019a).

Equally crucial is to recognise that multimorbidity is associated with social and economic determinants (Skivington et al., 2015) that can add to (or interact with) behavioural determinants. For example, advanced age, female gender, low socioeconomic status, and education have all been identified as significant risk factors for the onset of multimorbidity (Northwood et al., 2018). These findings align with the Social Determinants of Health (SDoH) framework, which describes how broader societal structures – from economic policies to social norms – shape and segment populations hierarchically based on gender, race, education, occupation, and income (Braveman and Gottlieb, 2014; Fuchs, 2017). This stratification then directly and indirectly influences health outcomes. The present research therefore incorporates insights from the SDoH framework to examine whether socio-demographic factors predict membership within risk behaviour clusters. Given that socio-demographic determinants can not only shape individual health behaviours but also predict health outcomes through

complicated multifactorial pathways, we will also adjust for them in examining the relationship between health behaviours and outcomes (Northwood et al., 2018).

The Present Research

The present research analyses data from a longitudinal panel of older adults in England to: i) explore how the SNAP behaviours cluster over time in older adults, ii) investigate how membership in different behavioural clusters varies by socio-demographic characteristics, and iii) examine which, if any, behavioural clusters are prospectively associated with multimorbidity over time.

Methods

Study design

We analysed secondary data from the English Longitudinal Study of Ageing (ELSA) – a nationally representative, ongoing panel study of community-dwelling adults aged 50 and over at baseline, in England (Banks, 2021). ELSA collects biennial data on mental and physical health, finances, and attitudes around ageing using computer-assisted interviews and questionnaires (Banks, 2021). This study followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (See Supplementary Section 5) (Von Elm et al., 2007). Ethical approval for ELSA was obtained from the National Research Ethics Service. All participants provided written informed consent. Separate ethical approval and consent were not required for our analyses because data were fully anonymised.

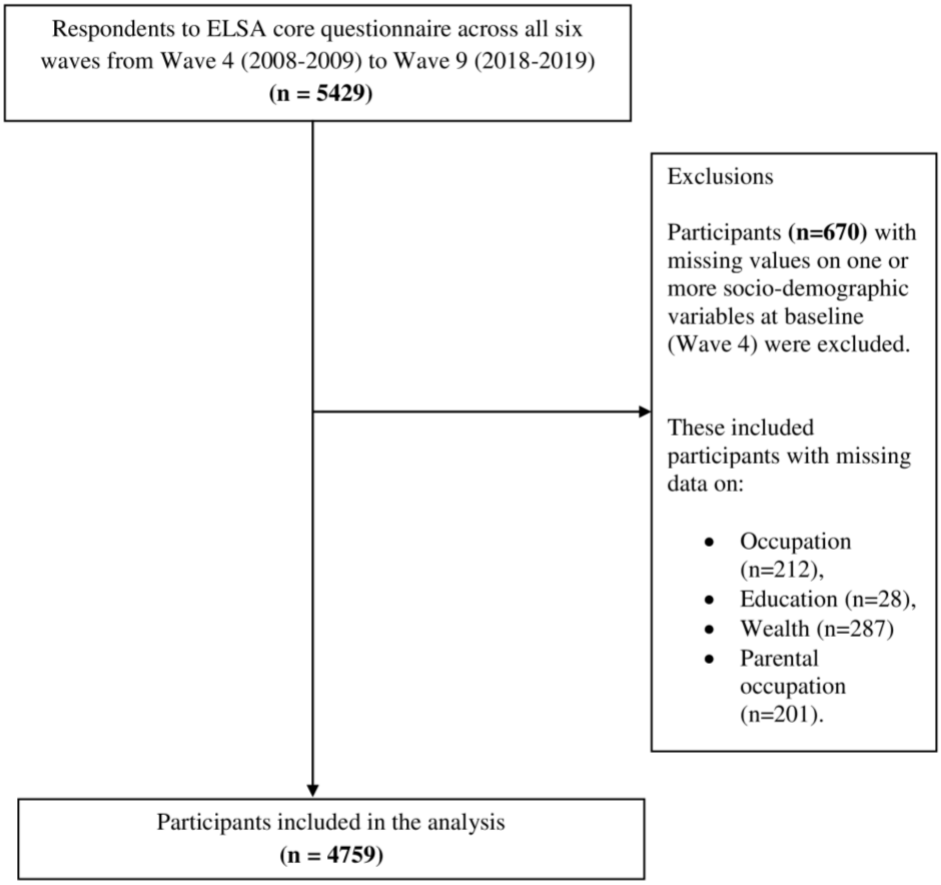
Sample Selection and Exclusion Criteria

Our analysis used data from 5,429 respondents to the core questionnaire across six waves from Wave 4 (2008-2009) to Wave 9 (2018-2019). We applied the longitudinal weights that were provided with the dataset and had been derived using information spanning from Wave 4 to Wave 9 to reduce drop-out bias. Wave 4 was selected as the baseline because, although data on health behaviours was available from Wave 3, longitudinal weights were only available from Wave 1 or Wave 4. Choosing Wave 3 as the baseline would have resulted in the loss of data on approximately 2000 participants due to longitudinal weighting.

Participants (n=670) with missing values on socio-demographic variables were removed using listwise deletion, leaving a final sample of 4759 participants (87.6% of the original sample; see Fig 1). We chose to use listwise deletion because the MPlus v8.5 software package does not support handling missing data for socio-demographic predictors in a latent class analysis (see

Figure 1). More specifically, data was missing for occupation (n=212), education (n=28), wealth (n=287), and parental occupation (n=201). Given that no socio-demographic variable had more than 5% missing data, a threshold below which multiple imputation is deemed less beneficial, we favoured a complete case analysis (Schafer, 1999).

Fig 1. Study flow chart



To assess the potential impact of excluding participants with missing data, we compared the included and excluded samples (for details, see Supplementary Section 2). Overall, the absolute differences between the included and excluded samples were not substantial, though differences for some socio-demographic variables (average age, tertiary education, intermediate and professional/managerial occupations) and disease status (complex multimorbidity, respiratory disorders and endocrine disorders) achieved significance due to the relatively large number of participants.

Measures of health behaviour

Data on health behaviours was taken from Waves 4-8. ELSA provides several measures of SNAP behaviours. Therefore, an online survey was used to gather consensus from experts on how SNAP behaviours should be *defined* for the present analysis (i.e. which measure to choose) and *categorised* into risk groups (i.e. how many categories to divide each behaviour into and what cut-offs to use; see Supplementary Section 1). We describe the agreed measures below.

Smoking

We used current smoking status as a measure of smoking. The data comprised participants' binary (yes/no) response to the question 'Do you smoke at all nowadays?'

Fruit and vegetable intake

In Wave 4, fruit and vegetable intake was assessed with questions such as 'How much of the following did you eat yesterday?' for 13 foodstuffs including a 'small glass of fruit juice' and 'salad (cereal bowlfuls)'. However, across Waves 5 to 8, participants were asked 'how many portions of vegetables – excluding potatoes – do you eat on a given day?' and 'how many portions of fruits do you eat on a given day?' To make fruit and vegetable intake consistent

across waves, we used an established method to match Wave 4 and Waves 5 to 8 responses (see Supplementary Fig S1)(Hackett et al., 2018; Kojima et al., 2020). Specifically, portions of fruit and vegetable consumed per day were added to create a single variable for each Wave, which was subsequently divided into two categories (<5 or ≥5 portions per day).

Alcohol Consumption

We included data on alcohol consumption over the last week: Specifically, i) whether the participant consumed alcohol in the last week (yes/no), and ii) the volume of each alcoholic beverage (i.e. beer, wine, and spirits) consumed. Consumption was converted into the number of UK units (1 UK unit = 8g of pure alcohol) using standard assumptions for beverage strength (Meier et al., 2021). Consumption was then categorised into four levels based on units consumed per week: *harmful* (>50 units for men, >35 units for women); *hazardous* (>14-50 units for men and >14-35 units for women); *moderate* (14 units or less); *and abstainers* (0 units) (Health, 2016).

Physical Activity

Physical activity was recorded in ELSA by asking participants how often they took part in each of three types of physical activity: vigorous-intensity (e.g. running/jogging, swimming, etc.), moderate-intensity (e.g. gardening, cleaning the car, etc.) and low-intensity (e.g. laundry and home repairs). The response categories were: hardly ever/never, one to three times a month, once a week and more than once a week. Following previous research (Dhalwani et al., 2016; Hamer, de Oliveira and Demakakos, 2014), we created a summary index by adding responses to all three questions and classified participants' levels of physical activity as: *sedentary* (no activity on a weekly basis); *low* (only mild activity at least once a week); *moderate* (moderate

but no vigorous activity at least once a week); or *high* (any vigorous activity at least once a week).

Socio-demographic variables

Socio-demographic variables were taken from Wave 4 and included: *age, sex, parental occupation, own occupation, education* and *wealth*. For *parental occupation* and participants' *own occupation,* the three-class version of the National Statistics—Socioeconomic Classification Scheme (Elias and McKnight, 2003) was used: professional and managerial occupations, intermediate occupations, and semi-routine and routine occupations. Participants with occupations listed as not classifiable (n=12) were excluded. *Education* was taken from the Wave 4 IFS derived dataset and grouped into 'degree/higher' (National Vocational Qualification NVQ4/NVQ5/degree or equivalent), 'intermediate' (higher education below degree, NVQ3/GCE A-level equivalent, NVQ2/GCE O-level equivalent, NVQ1/CSE other grade equivalent or foreign/other), 'no qualifications' (Fraser et al., 2014). *Wealth* was chosen as the most appropriate economic indicator for participants aged 50 and above, as it better reflects their financial resources during active professional life and retirement compared to income. The ELSA derived dataset defines wealth as the net total non-pension wealth including property, possessions, housing, investments, savings, artwork, jewellery, and net of debt reported at the household level (i.e. an individual or a couple living at the same address who make joint financial decisions)(Tsimpida et al., 2022). Wealth was grouped into tertiles to reduce measurement error and facilitate comparisons of health measures across equally sized groups within the population. All variables (except age) were converted into dummy variables, with the lowest category serving as the reference group.

Disease status

We assessed the disease status of 25 physical and mental health conditions recorded at each wave as listed in Table 1, to evaluate basic multimorbidity and complex multimorbidity (Singer et al., 2019a). Disease data was collected from the baseline Wave 4, and from the final Wave 9. Participants were asked whether they still had the condition diagnosed by a doctor that they had reported previously and, if not, whether they could report a new condition.

Researchers have measured multimorbidity using various methods (Wister et al., 2015). Given the lack of a uniform approach, we adopted the most widely cited and accepted definition of basic multimorbidity, which identifies it as having two or more chronic conditions (Mercer et al., 2016). Consequently, we coded respondents as "yes" for multimorbidity if they had two or more conditions from the 25-condition list and as "no" otherwise.

Complex multimorbidity was defined as having three or more conditions affecting three or more body systems (Harrison et al., 2014; Singer et al., 2019a). We based our selection of body system disorders (listed in Table 2) for calculating complex multimorbidity on a previous study using the ELSA dataset (Singer et al., 2019a). This study identified eight body systems as outlined by the International Classification of Diseases 10th Revision system (DiSantostefano, 2009): eye disorders; circulatory disorders; nervous disorders; mental and behavioural problems; neoplasms; respiratory disorders; endocrine, nutritional and metabolic disorders; and musculoskeletal and connective system disorders. Using the self-reported presence or absence of three or more body system disorders, we derived a binary variable representing complex multimorbidity.

Table 1. Morbidities used to ascertain multimorbidity and complex multimorbidity

Body system disorders	Morbidities
1. Eye disorders	1. Glaucoma
	2. Macular degeneration
	3. Cataracts
2. Circulatory disorders	1. High blood pressure
	2. Angina
	3. Heart Attack
	4.Congestive heart failure
	5. Heart murmur
	6. Abnormal heart rhythm
	7. Stroke
3. Endocrine, nutritional and metabolic	1. Diabetic eye disease
	2. Diabetes
4. Musculoskeletal and connective system	1. Osteoporosis
	2. Arthritis
5. Respiratory	1. Lung disease
	2. Asthma
6. Neoplasms	1. Cancers
7. Nervous disorders	1. Parkinson's disease
	2. Alzheimer's disease
	3. Hallucinations
8. Mental and behavioural	1. Anxiety
	2. Depression
	3. Emotional problems
	4. Mood swings
	5 Dementia

Note. Adapted from Singer et al. (Singer et al., 2019a)

Statistical analysis

RMLCA was used to examine whether there were distinct classes of respondents who had similar patterns of SNAP behaviours over time. RMLCA was chosen as it adopts a probabilistic model-based approach for capturing the number and composition of clusters, handles categorical data well, and allows for reliable interpretation and replication of patterns uncovered in the data. MPlus v8.5 software and R version v4.0.3 (Muthén and Muthén, 2017b; Nylund-Gibson and Choi, 2018) was used to conduct the RMLCA.

A two-stage approach was used. In the first stage, the optimal number of classes (i.e. clusters) was determined (Aim 1). For this, data on the four health behaviours across the five waves was entered as independent data points to create a series of LCA models with increasing numbers

of latent classes (i.e. clusters) until the model fit stopped improving. The fit of the models was evaluated using several indices: Consistent Akaike's Information Criterion (CAIC), Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (aBIC), Approximate Weight of Evidence Criterion (AWE), Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT). More details on the evaluation of model fit are provided in Supplementary Section 3. Missing data on the health behaviours were accounted for by Full Information Maximum Likelihood (Nylund-Gibson and Choi, 2018). To assess the reliability of the class solution, we conducted a split-half replication where the sample was randomly split in half, and the above RMLCA was performed separately on these split samples to see if the solution for the full sample was replicated between these smaller splits. (for details, see Supplementary Section 4). In the second stage, we applied the 3-step method proposed by Bolck, Croon, and Hagenaars (Nylund-Gibson and Choi, 2018). This method assigns individuals to the class that they have the highest posterior probability of belonging to.

To examine the association between socio-demographic characteristics and class membership (Aim 2), we regressed latent classes on socio-demographic variables in a series of multinomial logistic regressions, controlling for the potential inaccuracies in the class assignments, also known as classification errors, that are extracted as part of the 3-step method.

To assess whether the prevalence of each health condition differed across classes (Aim 3), we regressed each health outcome on the latent classes in a series of binomial logistic regressions, adjusting for: i) socio-demographic variables, ii) respective disease at baseline, and iii) classification errors extracted as part of the 3-step method that account for uncertainty in class assignment. Then, for each health outcome, an omnibus Wald chi-square test was conducted to test for differences in disease prevalence across all classes (α=0.05). If significant, pairwise

comparisons were conducted to test for differences in disease prevalence between each pair of classes. The significance level for pairwise comparisons was adjusted using the Bonferroni correction ($\alpha = 0.007$).

Results

Sample characteristics

On average, participants were aged 62.9 years (SD = 8.1) and approximately half were female (56.3%; see Table 2 for baseline demographic data). The sample's engagement in health behaviours across waves is shown in Table 3. The body system disorders with the highest prevalence were: multimorbidity (57.9%), circulatory disorders (50.3%), and disorders of the musculoskeletal and connective system (47.5%) (see Supplementary Table 3).

Table 2. Baseline socio-demographic data of the final sample (n=4759)

Baseline socio-demographic characteristics	**Included sample (n=4759) Mean (SD)**
Average Age	62.9 (8.1)
	N (%)
Male	2081 (43.7)
Female	2678 (56.3)
Parental Occupation – Semi-routine and routine	1299 (27.3)
Parental Occupation – Intermediate	1530 (32.1)
Parental Occupation – Professional/managerial	1930 (40.6)
Occupation – Semi-routine and routine	1646 (34.6)
Occupation – Intermediate	1245 (26.2)
Occupation – Professional/managerial	1868 (39.3)
Education – No qualifications	958 (20.1)
Education – Intermediate	2807 (59.0)
Education – Degree/higher	994 (20.9)
Wealth – First Tertile	1260 (26.5)
Wealth – Second Tertile	1615 (33.9)
Wealth – Third Tertile	1884 (39.6)

Table 3. Class-defining indicators (i.e. SNAP behaviours) among participants (n= 4759) included in the latent class analysis

		Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
		N (%)	N (%)	N (%)	N (%)	N (%)
Smoking	Non-smoker	3733 (78.4)	2321 (48.8)	2471 (51.9)	2555 (53.7)	2625 (55.2)
	Smoker	525 (11)	484 (10.2)	432 (9.1)	399 (8.4)	353 (7.4)
	Missing	501 (10.5)	1954 (41.1)	1856 (39)	1805 (37.9)	1781 (37.4)
Alcohol consumption	Abstainer	789 (16.6)	929 (19.5)	934 (19.6)	920 (19.3)	961 (20.2)
	Moderate	1780 (37.4)	1763 (37)	1702 (35.8)	1700 (35.7)	1628 (34.2)
	Hazardous	1092 (22.9)	1111 (23.3)	1032 (21.7)	980 (20.6)	966 (20.3)
	Harmful	183 (3.8)	176 (3.7)	177 (3.7)	165 (3.5)	150 (3.2)
	Missing	915 (19.2)	780 (16.4)	914 (19.2)	994 (20.9)	1054 (22.1)
Physical activity	Sedentary	22 (0.5)	22 (0.5)	29 (0.6)	27 (0.6)	30 (0.6)
	Low	2345 (49.3)	2389 (50.2)	2466 (51.8)	2578 (54.2)	2658 (55.9)
	Moderate	804 (16.9)	791 (16.6)	762 (16)	767 (16.1)	680 (14.3)
	High	1121 (23.6)	1081 (22.7)	1077 (22.6)	975 (20.5)	914 (19.2)
	Missing	467 (9.8)	476 (10)	425 (8.9)	412 (8.7)	477 (10)
Fruit and vegetable intake	< 5 portions/day	1761 (37)	1820 (38.2)	1693 (35.6)	1553 (32.6)	1645 (34.6)
	>= 5 portions/day	2540 (53.4)	2656 (55.8)	2706 (56.9)	2753 (57.8)	2683 (56.4)

Missing	458 (9.6)	283 (5.9)	360 (7.6)	453 (9.5)	431 (9.1)

Health behaviour clusters and within-cluster changes

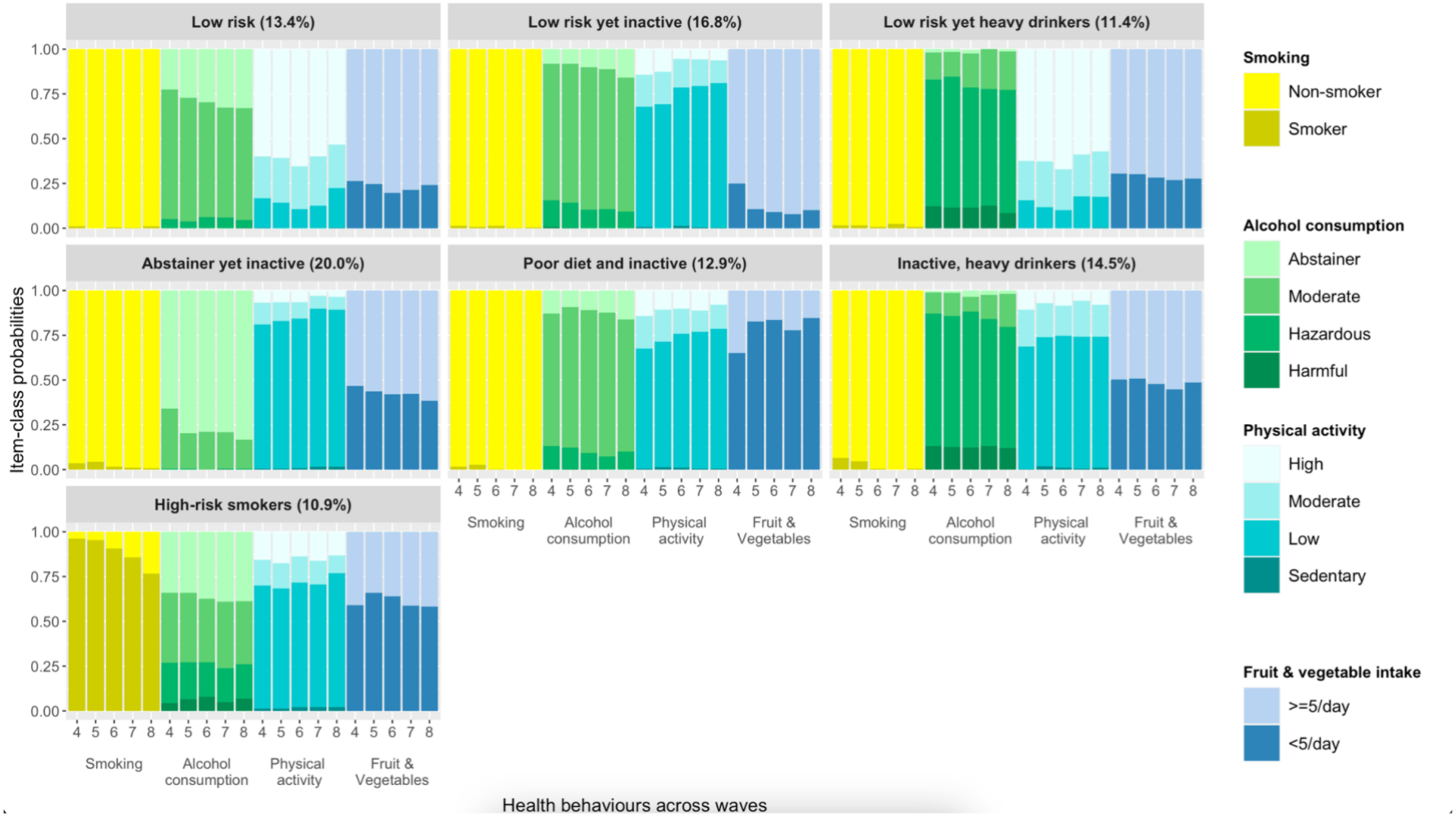
After fitting models with one to nine latent classes, a model with seven classes was considered the best fit (model fit statistics are shown in Table 4). The VLMR-LMR test supported a five- and seven-class solution. However, the values of the BIC, SABIC, and CAIC continued to decrease as the number of classes increased, suggesting improved model fit (Chen et al., 2017). As the decline in information criteria plateaued around seven classes (screeplot shown in Supplementary Fig S2), we opted for the seven-class model. The entropy and smallest average latent class posterior probability fell within the recommended range (value ≥0.8) for the seven-class model (Nylund-Gibson and Choi, 2018). Furthermore, analysis on the split-half samples also indicated that the 7-class solution was the optimal solution, as identical classes were uncovered in each split-half (full details can be found in Supplementary Section 3). The seven distinct health-risk behaviour classes — hereafter referred to as clusters — were assigned labels. Fig 2 depicts these clusters and the average probability that individuals in each cluster engaged in the four health behaviours—smoking, alcohol consumption, physical activity, and fruit and vegetable intake—over five distinct time points. Most behaviours were fairly stable over time, so unless highlighted below the behaviours that characterise each cluster were similar at each time point.

Table 4. Model fit evaluation information for choosing a latent class model

K	LL	CAIC	BIC	SABIC	AWE	VLMR LRT p-value	Entropy	Smallest average latent class posterior probability	Smallest class size (%)
1	-64393.26	128866.51	129125.23	128998.12	128796.5	-	-	1	100
2	-60051.70	120265.39	120789.28	120531.90	120123.3	<0.001	0.810	0.943	42.5
3	-56767.60	113779.20	114568.27	114180.60	113565.1	<0.001	0.870	0.936	12.4
4	-55061.43	110448.86	111503.11	110985.15	110162.7	<0.001	0.860	0.892	11.6
5	-54186.53	108781.06	110100.49	109452.25	108422.9	<0.001	0.830	0.823	11.2
6	-53544.39	107578.79	109163.40	108384.88	107148.6	0.770	0.830	0.819	11
7	**-52947.80**	**106467.60**	**108317.39**	**107408.58**	**105965.4**	**<0.001**	**0.820**	0.820	10.9
8	-52464.47	105582.94	107697.91	106658.82	105008.7	0.760	0.820	0.819	3
9	-52081.57	104899.14	107279.29	106109.92	104252.9	0.760	0.820	0.810	3.2

Note: n=4759; K = number of classes (the nine-class model failed to converge); LL = model log likelihood; BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; p-value significance <0.05.

Fig 2. Seven-class model reflecting different clusters of health behaviour across time.



Note. The x-axis lists each of the four behaviours – smoking, alcohol consumption, physical activity, and fruit and vegetable intake – across five time points. The y-axis provides the average probability for each of the indicators (i.e. four health behaviours) conditional on membership in a given class (i.e. cluster).

Participants in the *Low-risk* cluster (13.4% of the sample) maintained high levels of physical activity and adequate (≥5 per day) fruit and vegetable intake across time. They also had a consistently high probability of drinking at moderate levels, although the proportion of people who abstain from alcohol rose across waves.

Participants in the *Low-risk yet inactive* cluster (16.8%) displayed low levels of physical activity. However, they consumed the recommended portions of fruit and vegetables, and had moderate alcohol intake.

Participants in the *Low-risk yet heavy drinkers* cluster (11.4%) exhibited patterns of smoking, physical activity, and diet similar to those in the *Low-risk* cluster. However, participants in the *Low-risk yet heavy drinkers* cluster were more likely to drink at hazardous and harmful levels.

The largest proportion of participants fell in the *Abstainer yet inactive* cluster (20%). Although participants in this cluster displayed low levels of physical activity, they were also the least likely to consume alcohol at harmful or hazardous levels and had the highest probability of abstaining from alcohol which steadily rose across waves.

Participants in the *Poor diet and inactive* cluster (12.9%) had a similar profile to the *Low-risk yet inactive* cluster, except they had a consistently low probability of adequate fruit and vegetable intake — the lowest of any cluster.

Participants in the *Inactive, heavy drinkers* cluster (14.5%) had a high probability of drinking at hazardous and harmful levels while displaying consistently low levels of physical activity.

Finally, participants characterised as *High-risk smokers* cluster (10.9%) had a high probability of smoking, albeit declining over time, low levels of physical activity, and inadequate fruit and vegetable intake.

Clusters and socio-demographic characteristics

Table 5 displays the socio-demographic composition of participants in each of the identified clusters and the results from the adjusted multinomial logistic regressions that examined associations between each socio-demographic characteristic and cluster membership.

Compared to participants in the *Low-risk* cluster (which served as the reference group), participants characterised as *Low risk yet inactive*, *Abstainers but inactive*, *Poor diet and inactive*, and *Inactive, heavy drinkers* were more likely to be older. In contrast, individuals in the *High-risk smokers* cluster were more likely to be younger than participants in the *Low-risk* cluster. However, while these age-related associations are notable, they were weaker than the associations found with other socio-demographic factors.

Table 5. Demographics and Odds ratios from Multinomial Logistic regressions examining the association between socio-demographic predictors and cluster membership.

Socio-demographic characteristics	Low risk (n=13.4 %) (Ref. class)		Low risk yet inactive (n=16.8%)		Low risk yet heavy drinkers (n=11.4%)		Abstainers but inactive (n=20 %)		Poor diet and inactive (n=12.9 %)		Inactive, heavy drinkers (n=14.5 %)		High-risk smokers (n=10.9 %)	
				OR [95% C.I.]		OR [95% C.I.]		OR [95% C.I.]		OR [95% C.I.]		OR [95% C.I.]		OR [95% C.I.]
Age	61.42	Ref.	65.30	**1.06 [1.04, 1.08]**	60.31	0.97 [0.96, 1.00]	66.70	**1.07 [1.05, 1.09]**	65.00	**1.06 [1.03, 1.08]**	62.97	**1.03 [1.01, 1.05]**	60.52	**0.97 [0.95, 0.99]**
(s.d.)	(8.4)		(12)		(7.7)		(13.2)		(13.5)		(11.3)		(8.7)	
Sex														
Male	45.6%	Ref	35.5%	Ref	67.5%	Ref	25.4%	Ref	51.6%	Ref	69.1%	Ref	45.2%	Ref
Female	54.4%	Ref	64.5%	**1.49 [1.10, 2.02]**	32.5%	**0.40 [0.29, 0.55]**	74.6%	**2.31 [1.68, 3.17]**	48.4%	0.77 [0.55, 1.06]	30.9%	**0.37 [0.27, 0.49]**	54.8%	1.02 [0.75, 1.40]
Education Level														
No qualifications	15.5%	Ref	23.4%	Ref	11.3%	Ref	43.9%	Ref	30.1%	Ref	13.4%	Ref	40.5%	Ref
Intermediate	58.1%	Ref	61.4%	0.89 [0.57, 1.39]	52.9%	0.90 [0.53, 1.53]	50.6%	**0.56 [0.38, 0.83]**	60.9%	0.76 [0.49, 1.18]	62.7%	1.24 [0.78, 1.96]	51.2%	**0.44 [0.29, 0.66]**
Degree or higher	26.4%	Ref	15.2%	**0.52 [0.30, 0.88]**	35.8%	0.91 [0.51, 1.63]	5.5%	**0.23 [0.13, 0.40]**	9.0%	**0.32 [0.18, 0.60]**	23.9%	0.84 [0.50, 1.42]	8.3%	**0.21 [0.12, 0.36]**
Wealth														
First tertile	15.8%	Ref	25.0%	Ref	9.5%	Ref	47.8%	Ref	37.2%	Ref	20.6%	Ref	50.9%	Ref
Second tertile	35.5%	Ref	37.2%	0.67 [0.43, 1.03]	27.9%	1.17 [0.67, 2.06]	33.9%	**0.38 [0.26, 0.57]**	41.2%	**0.53 [0.34, 0.81]**	30.3%	**0.63 [0.40, 0.97]**	30.5%	**0.33 [0.22, 0.49]**
Third tertile	48.7%	Ref	37.8%	**0.48 [0.31, 0.75]**	62.6%	1.71 [0.99, 2.94]	18.3%	**0.18 [0.12, 0.28]**	21.6%	**0.22 [0.14, 0.36]**	49.1%	0.71 [0.47, 1.09]	18.6%	**0.18 [0.11, 0.28]**
Occupation - Self														
Routine/manual	33.3%	Ref	36.8%	Ref	18.5%	Ref	55.8%	Ref	45.8%	Ref	31.4%	Ref	54.1%	Ref
Intermediate	27.0%	Ref	27.7%	1.11 [0.75, 1.64]	26.3%	**1.70 [1.07, 2.71]**	22.6%	0.84 [0.57, 1.22]	28.2%	1.17 [0.77, 1.76]	22.5%	1.03 [0.69, 1.52]	21.9%	0.87 [0.58, 1.30]
Professional/managerial	39.7%	Ref	35.5%	1.32 [0.90, 1.94]	55.2%	**1.95 [1.26, 3.04]**	21.6%	1.02 [0.70, 1.49]	26.0%	1.06 [0.70, 1.62]	46.1%	1.33 [0.93, 1.91]	24.0%	0.96 [0.65, 1.43]
Parental Occupation														
Routine/manual	24.2%	Ref	27.3%	Ref	20.8%	Ref	37.5%	Ref	29.5%	Ref	25.1%	Ref	35.9%	Ref
Intermediate	35.0%	Ref	28.8%	0. 78 [0.53, 1.15]	29.7%	0.82 [0.53, 1.25]	34.1%	0.79 [0.54, 1.14]	38.5%	1.06 [0.71, 1.60]	28.9%	0.77 [0.52, 1.13]	40.2%	0.96 [0.66, 1.40]

Professional/managerial	40.8%	Ref	43.9%	1.14 [0.78, 1.67]	49.5%	1.11 [0.73, 1.67]	28.4%	0.85 [0.58, 1.23]	32.0%	1.10 [0.71, 1.71]	46.0%	1.18 [0.81, 1.71]	23.9%	0.76 [0.51, 1.14]

Note. Odds Ratios [95% Confidence interval] are from BCH multinomial logistic regression analysis; Ref = Reference cluster. **Bold values** are statistically significant at the significance level (p=0.05)**.** All clusters are compared to the Reference cluster - *Low-risk*. Each odds ratio is adjusted for the remaining socio-demographic variables in the model.

Men made up around 70% of two clusters: *Low-risk yet heavy drinkers* and *Inactive, heavy drinkers*. By contrast, the two clusters labelled *Low-risk yet inactive* and *Abstainers but inactive* were predominantly female, with the likelihood of female membership in these groups being 1.5 to 2 times higher than the *Low-risk* cluster (reference group).

Notably, the *Low-risk yet inactive* and *Abstainers but inactive* clusters were less likely to comprise individuals with degree-level education or those belonging to higher wealth tertiles, compared to the *Low-risk* cluster. Similarly, participants characterised as *High-risk smokers* and *Poor diet and inactive* were also less likely to belong to the higher tiers of both wealth and education compared to participants in the *Low-risk* cluster.

Individuals characterised as *Low-risk yet heavy drinkers* were more likely to have professional or managerial occupations than those in the *Low-risk* cluster.

Health behaviour clusters and disease status

The prevalence of multimorbidity, complex multimorbidity, respiratory disorders, and endocrine, nutritional, and metabolic disorders differed significantly across clusters (omnibus Wald test χ^2(df = 6) multimorbidity = 14.954 , p = 0.021; χ^2(df) = 6 complex multimorbidity = 31.326, p <0.001; χ^2(df = 6) respiratory disorders = 35.998, p <0.001; and χ^2(df = 6) endocrine, nutritional and metabolic disorders = 53.201, p <0.001, respectively). To further investigate which clusters differed in the proportions of participants with each disease profile, we conducted pairwise comparisons between clusters for respiratory diseases, multimorbidity, complex multimorbidity, and endocrine, nutritional, and metabolic disorders, as shown in Table 6. Results were adjusted for disease status and socio-demographic variables at baseline Wave 4 (see Table 6; for unadjusted results, see Supplementary Table S5).

Participants characterised as *Low-risk* had a lower prevalence of multimorbidity compared to participants in the *Low risk yet inactive, Abstainers but inactive,* and *Inactive, heavy drinkers* clusters. Participants in the *Low-risk* cluster also had a lower prevalence of complex multimorbidity and endocrine disorders than participants in the *Low-risk yet inactive* and *Abstainers but inactive* clusters. Similarly, the *Low-risk yet heavy drinkers* had a lower prevalence of complex multimorbidity than all other clusters except two, namely, *Low-risk* and *Poor diet and inactive* clusters.

By contrast, *High-risk smokers* had a higher prevalence of respiratory disorders than all other clusters. Individuals in this cluster also a higher prevalence of complex multimorbidity compared to the *Low-risk* and *Low-risk yet heavy drinkers* clusters.

With regards to *endocrine, nutritional, and metabolic disorders,* participants in the *Low-risk yet inactive* cluster had a higher prevalence than the two clusters characterised by heavy drinking. However, the *Abstainers but inactive* cluster had a higher prevalence of endocrine disorders than all other clusters, except the *Low-risk yet inactive* cluster. *High-risk smokers* also had a higher prevalence of endocrine disorders than the *Low-risk yet heavy drinkers.*

Finally, the *Inactive, heavy drinkers* had a higher prevalence of multimorbidity than the *Low-risk* cluster and a higher prevalence of complex multimorbidity than the *Low-risk yet heavy drinkers.*

Table 6. Pairwise comparisons of disease status prevalence across clusters (adjusted for i) the specific disease at baseline Wave 4, and ii) socio-demographic variables – sex, age, parental occupation, own occupation, education level and wealth)

1

Health conditions			1	2	3	4	5	6	7
			Low risk (Ref. class)	**Low risk yet inactive**	**Low risk yet heavy drinkers**	**Abstainers but inactive**	**Poor diet and inactive**	**Inactive, heavy drinkers**	**High-risk smokers**
			(n=13.4 %)	(n=16.8%)	(n=11.4%)	(n=20 %)	(n=12.9 %)	(n=14.5 %)	(n=10.9 %)
Multimorbidity	1	0.469		-0.179	0.015	-0.242	-0.128	-0.125	-0.108
				(0.007)	(0.279)	**(0.003)**	(0.300)	**(0.005)**	(0.096)
	2	0.648	**0.179**		0.194	-0.063	0.051	0.054	0.071
			(0.007)		(0.094)	(0.760)	(0.115)	(0.983)	(0.320)
	3	0.454	-0.015	-0.194		-0.257	-0.143	-0.14	-0.123
			(0.279)	(0.094)		(0.053)	(0.999)	(0.114)	(0.549)
	4	0.711	**0.242**	0.063	0.257		0.114	0.117	0.134
			(0.003)	(0.760)	(0.053)		(0.073)	(0.745)	(0.202)
	5	0.597	0.128	-0.051	0.143	-0.114		0.003	0.02
			(0.300)	(0.115)	(0.999)	(0.073)		(0.130)	(0.560)
	6	0.594	**0.125**	-0.054	0.14	-0.117	-0.003		0.017
			(0.005)	(0.983)	(0.114)	(0.745)	(0.130)		(0.344)
	7	0.577	0.108	-0.071	0.123	-0.134	-0.02	-0.017	
			(0.096)	(0.320)	(0.549)	(0.202)	(0.560)	(0.344)	
Complex Multimorbidity	1	0.18		**-0.148**	0.057	**-0.226**	-0.09	-0.085	**-0.109**
				(0.007)	(0.287)	**(<0.001)**	(0.352)	(0.009)	**(0.006)**
	2	0.328	**0.148**		**0.205**	-0.078	0.058	0.063	0.039

			(0.007)		**(<0.001)**	(0.350)	(0.100)	(0.836)	(0.869)
	3	0.123	-0.057	**-0.205**		**-0.283**	-0.147	**-0.142**	**-0.166**
			(0.287)	**(<0.001)**		**(<0.001)**	(0.057)	**(0.001)**	**(<0.001)**
	4	0.406	**0.226**	0.078	**0.283**		0.136	0.141	0.117
			(<0.001)	(0.350)	**(<0.001)**		(0.014)	(0.263)	(0.471)
	5	0.27	0.09	-0.058	0.147	-0.136		0.005	-0.019
			(0.352)	(0.100)	(0.057)	(0.014)		(0.158)	(0.087)
	6	0.265	0.085	-0.063	**0.142**	-0.141	-0.005		-0.024
			(0.009)	(0.836)	**(0.001)**	(0.263)	(0.158)		(0.727)
	7	0.289	**0.109**	-0.039	**0.166**	-0.117	0.019	0.024	
			(0.006)	(0.869)	**(<0.001)**	(0.471)	(0.087)	(0.727)	
Respiratory disorders	1	0.096		-0.051	-0.003	-0.082	-0.018	-0.032	**-0.133**
				(0.096)	(0.415)	(0.016)	(0.620)	(0.234)	**(<0.001)**
	2	0.147	0.051		0.048	-0.031	0.033	0.019	**-0.082**
			(0.096)		(0.451)	(0.381)	(0.271)	(0.636)	**(<0.001)**
	3	0.099	0.003	-0.048		-0.079	-0.015	-0.029	**-0.13**
			(0.415)	(0.451)		(0.137)	(0.752)	(0.789)	**(<0.001)**
	4	0.178	0.082	0.031	0.079		0.064	0.05	**-0.051**
			(0.016)	(0.381)	(0.137)		(0.069)	(0.219)	**(0.003)**
	5	0.114	0.018	-0.033	0.015	-0.064		-0.014	**-0.115**
			(0.620)	(0.271)	(0.752)	(0.069)		(0.555)	**(<0.001)**
	6	0.128	0.032	-0.019	0.029	-0.05	0.014		**-0.101**
			0.234	0.636	0.789	0.219	0.555		**(<0.001)**
	7	0.229	**0.133**	**0.082**	**0.13**	**0.051**	**0.115**	**0.101**	
			(<0.001)	**(<0.001)**	**(<0.001)**	**(0.003)**	**(<0.001)**	**(<0.001)**	
Endocrine, nutritional and	1	0.083		**-0.08**	0.03	**-0.153**	-0.061	-0.016	-0.058
				(0.005)	(0.129)	**(<0.001)**	(0.294)	(0.633)	(0.104)

metabolic	2	0.163	**0.08**		**0.11**	-0.073	0.019	**0.064**	0.022
disorders			**(0.005)**		**(<0.001)**	(0.140)	(0.106)	**(<0.001)**	(0.188)
	3	0.053	-0.03	**-0.11**		**-0.183**	-0.091	-0.046	**-0.088**
			(0.129)	**(<0.001)**		**(<0.001)**	(0.012)	(0.218)	**(0.002)**
	4	0.236	**0.153**	0.073	**0.183**		**0.092**	**0.137**	**0.095**
			(<0.001)	(0.140)	**(<0.001)**		**(0.006)**	**(<0.001)**	**(0.006)**
	5	0.144	0.061	-0.019	0.091	**-0.092**		0.045	0.003
			(0.294)	(0.106)	(0.012)	**(0.006)**		(0.108)	(0.644)
	6	0.099	0.016	**-0.064**	0.046	**-0.137**	-0.045		-0.042
			(0.633)	**(<0.001)**	(0.218)	**(<0.001)**	(0.108)		(0.019)
	7	0.141	0.058	-0.022	**0.088**	**-0.095**	-0.003	0.042	
			(0.104)	(0.188)	**(0.002)**	**(0.006)**	(0.644)	(0.019)	

Note. The estimates are the absolute differences in proportions of participants having the disease in the cluster (in row) minus the cluster (in column). P-values are shown in brackets. **Bold values** are statistically significant at the Bonferroni-corrected significance level (p=0.007) and indicate the two-tailed p-values for pairwise Wald test for differences in disease proportion for the cluster (in row) minus the cluster (in column).

Discussion

The present research investigated the relationship between clusters of health-risk behaviours over time and multimorbidity in older adults. We identified seven distinct clusters of behaviour that resemble those found in previous studies from Germany (Schneider et al., 2009), Australia (Griffin et al., 2014), and Taiwan (Hsu et al., 2013) including a cluster characterised by an overall low level of risk, a cluster characterised by physical inactivity, and a cluster characterised by heavy alcohol consumption, non-smoking and low physical activity. Similarly, with the exception of a study focusing on Taiwanese men (Hsu et al., 2013), where the smokers were split across two clusters because of a relatively high prevalence of smoking, the smallest subgroup in each study comprised smokers who exhibited two or more risky behaviours, which parallels our finding that the *High risk smokers* represented the smallest cluster (~11% of the sample). However, our clusters diverged from the findings of a study focusing on six international ageing cohorts, likely because their study: excluded dietary data, included social activity as a behaviour, and used different measures for physical activity, alcohol consumption, and smoking (Liao et al., 2019).

The present research moved beyond existing research, however, by using longitudinal data to not only examine whether distinct clusters of health behaviours are found in older adults but also whether and how patterns of behaviour within each cluster change over time. We found that patterns of behaviour within the clusters were largely stable over time, with two exceptions: The proportion of current smokers steadily declined in the *High-risk smokers* cluster, while the proportion of alcohol abstainers gradually increased in clusters characterised by moderate or no alcohol consumption (i.e. the clusters labelled *Low-risk* and *Abstainer yet inactive*). Notwithstanding these exceptions, our findings support the idea that SNAP behaviours in older people are fairly stable and likely reflect lifelong habits (Botoseneanu and

Liang, 2012), emphasising the importance of addressing risk behaviours early in the life course to prevent negative health outcomes (Tinner et al., 2021). Additionally, the finding that behavioural patterns are relatively stable over time suggests that clustering in older adults can be accurately captured by cross-sectional studies.

The clusters also had different socio-demographic profiles. Consistent with alcohol consumption patterns in the UK (Meier et al., 2021), the two clusters of heavy drinkers were predominantly male. The clusters characterised by physical inactivity but no other risky behaviours (i.e. the *Low-risk yet inactive* and *Abstainer yet inactive* clusters) were primarily female, similar to findings in previous studies (Griffin et al., 2014; Schneider et al., 2009; Hsu et al., 2013; Liao et al., 2019). *High-risk smokers* were younger on average and, in contrast to previous research, we did not find evidence that high-risk smokers more likely to be men (Barta, Powell and Wisnivesky, 2019). This may be due to survivorship bias, as smoking is the leading cause of lung cancer deaths, but lung cancer occurs less frequently and has a better prognosis in women (Barta, Powell and Wisnivesky, 2019). We also found a marked consistency with previous studies looking at clusters of health behaviour among older adults (Schneider et al., 2009; Griffin et al., 2014; Hsu et al., 2013), in that we found that clusters characterised by physical inactivity (in combination with other risky behaviours) were less likely to be wealthy or well-educated, suggesting a link between socio-demographic inequalities and health behaviour clustering.

Importantly, identified clusters also differed in their disease status. Participants characterised as *Abstainers but inactive* and *Low-risk yet inactive* had a higher prevalence of complex multimorbidity and endocrine disorders than other low-risk clusters that engaged in health-promoting behaviours (i.e. *Low-risk* and *Low-risk yet heavy drinkers*), and they also had higher

rates of multimorbidity compared to the *Low-risk* cluster. Notably, participants in the cluster characterised as *Abstainers but inactive* had a higher prevalence of endocrine disorders than participants in all clusters except *Low-risk yet inactive.* That the cluster characterised by physical inactivity (but no other risk behaviours) was associated with worse health outcomes than clusters characterised by multiple risk behaviours suggests that the relationship between behaviour and health outcomes is more complex than a linear dose–response relationship (Myint et al., 2011). Indeed, it is important to recognise the possibility of a bidirectional relationship between physical activity and multimorbidity, since not only is physical activity a risk factor for multimorbidity, but multimorbidity, in turn, can reduce function and reduce adherence to recommended levels of physical activity (Salman and Sellami, 2019).

Some associations were more straightforward and predictable. For example, we found that *High-risk smokers* had higher rates of respiratory disorders than every other cluster as might be expected. However, *High-risk smokers* also had a higher prevalence of complex multimorbidity and endocrine, nutritional and metabolic disorders compared to the *Low-risk yet heavy drinkers* cluster, a finding that is harder to explain using health behaviour patterns alone. Similarly hard to explain is the finding that *Inactive, heavy drinkers* had a higher prevalence of complex multimorbidity than *Low-risk yet heavy drinkers.* One explanation for the lower prevalence of complex multimorbidity and endocrine, nutritional and metabolic disorders could be that, compared to other clusters, the *Low-risk yet heavy drinkers* cluster had the largest proportion of individuals in the highest wealth tertile and in intermediate and professional jobs – indicators of elevated socioeconomic status. This higher socioeconomic status, a known protective factor, may influence health outcomes, as it has consistently been identified as an important determinant of multimorbidity [23]. Thus, examining the interaction between health behaviour clusters and socio-demographic variables on multimorbidity, could

further help clarify the patterns of risk. Additionally, our focus on the adverse health effects of risky behaviours might have overshadowed the protective effects of engaging in some behaviours (i.e., adequate fruit and vegetable intake and being physically active) [46]. Recognizing the potential benefits of these behaviours and their associated factors is crucial, as they offer functional, social, and psychological resilience against multimorbidity (Wister et al., 2022).

Strengths and limitations

Several strengths distinguish this study. It is the first to examine the association between longitudinal clusters of multiple health-risk behaviours and multimorbidity in older adults. Health behaviour experts helped to choose the most viable of the measures available in ELSA and how these might be used. It uses a robust, model-based, probabilistic approach (namely, RMLCA), demonstrates stable results in split-half replication, is reproducible (i.e. diagnostic criteria and programming codes are accessible) (Suhag, 2023). Furthermore, the results adjust for baseline disease and a range of socio-demographic variables that may confound the relationship between health behaviour and outcomes.

Despite these strengths, the study has some limitations. ELSA relies on self-report data, which can be subject to recall limitations and social desirability bias. Having said this, longitudinal analyses are less susceptible to misclassification bias due to consistent measures across survey waves. It is also important to note that alcohol consumption was only measured for the past week, which may misestimate drinking behaviour for those with inconsistent drinking patterns. Relatedly, missing data are unavoidable in general population cohorts such as ELSA and we had to exclude participants with missing sociodemographic data at baseline. As a result, participants who were included were slightly older, better educated, and more likely to have

more intermediate and professional level jobs than those who were excluded. This may limit the generalisability of our findings. Finally, as there are relatively few ethnic minority participants in ELSA, the findings may not generalise to non-white populations.

Implications

The present research offers new insights into the relationship between clusters of health behaviours and multimorbidity in older adults and has practical implications for interventions to improve health outcomes. For instance, by identifying distinct profiles of risk behaviour, our findings can help to identify high-risk subgroups and select behaviour(s) to target with interventions. For example, our data suggest that targeting physical inactivity, which characterised all five clusters associated with negative health outcomes and represented the majority of the population (70%), could have the greatest potential reach and health impact.

The present findings also demonstrate how targeting different clusters may require tailored approaches. For instance, interventions targeting participants in the *Abstainer yet inactive* cluster, which comprised mostly women and had lower levels of education and wealth, may need to address barriers to physical activity that are specific to their socio-demographic profile. This aligns with existing evidence indicating that interventions tailored to specific target audiences are more effective in promoting changes in multiple behaviours in the general adult population (Wilson et al., 2015; Sisti et al., 2018) and in patients with chronic conditions (Bricca et al., 2023), than interventions that are not tailored.

Conclusions

The present research identified seven clusters of older adults with distinct patterns of behaviour that were associated with socio-demographic characteristics and the prevalence of

multimorbidity. Notably, we found that the number or combination of risk behaviours alone could not explain why some clusters had worse health outcomes than others. A closer examination of how behaviour clusters interact with socio-demographic characteristics could offer a more nuanced understanding of their combined effect on health outcomes. Integrating this additional layer of complexity into our current study would have made its breadth unmanageable, but it remains an important area for future investigations to explore. Additionally, our findings show that health-risk behaviours tend to be stable as people age, emphasising the importance of addressing them early. Future research should take a lifespan approach to investigate how risk behaviours cluster at earlier life stages.

Declaration of Interests

We declare no competing interests.

Contributors

JH, TLW, and AS conceptualised and designed the study. AS conducted the statistical analyses under the supervision of TLW and JH. AS wrote the first draft and all authors provided critical revisions and approved the final submitted version.

Acknowledgements

We would like to thank Nicola Buckland, Robert Copeland, and Matt Field for their contributions to helping identify which measures from ELSA to focus on.

Ethical approval

ELSA received ethical approval from the National Research Ethics Service and all participants provided informed consent. Separate ethical approval and consent were not required for our analyses because data were fully anonymised.

Data sharing

The raw data involved in this analysis are available through the UK Data Service.

References

1. Kingston A, Robinson L, Booth H, Knapp M, Jagger C, Project M. Projections of multi-morbidity in the older population in England to 2035: estimates from the Population Ageing and Care Simulation (PACSim) model. Age and ageing. 2018;47(3):374-80.
2. Canizares M, Hogg-Johnson S, Gignac MA, Glazier RH, Badley EM. Increasing trajectories of multimorbidity over time: birth cohort differences and the role of changes in obesity and income. The Journals of Gerontology: Series B. 2018;73(7):1303-14.
3. Picco L, Achilla E, Abdin E, Chong SA, Vaingankar JA, McCrone P, et al. Economic burden of multimorbidity among older adults: impact on healthcare and societal costs. BMC health services research. 2016;16:1-12.
4. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet. 2012;380(9836):37-43.
5. Collaborators GRF. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet (London, England). 2017;390(10100):1345.
6. Nigg CR, Long CR. A systematic review of single health behavior change interventions vs. multiple health behavior change interventions among older adults. Translational behavioral medicine. 2012;2(2):163-79.
7. Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. Preventive medicine. 2015;81:16-41.
8. Botoseneanu A, Liang J. The effect of stability and change in health behaviors on trajectories of body mass index in older Americans: a 14-year longitudinal study. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences. 2012;67(10):1075-84.
9. Tamakoshi A, Tamakoshi K, Lin Y, Yagyu K, Kikuchi S, Group JS. Healthy lifestyle and preventable death: findings from the Japan Collaborative Cohort (JACC) Study. Preventive medicine. 2009;48(5):486-92.
10. McAloney K, Graham H, Law C, Platt L. A scoping review of statistical approaches to the analysis of multiple health-related behaviours. Preventive medicine. 2013;56(6):365-71.
11. Katikireddi SV, Skivington K, Leyland AH, Hunt K, Mercer SW. The contribution of risk factors to socioeconomic inequalities in multimorbidity across the lifecourse: a longitudinal analysis of the Twenty-07 cohort. BMC medicine. 2017;15(1):1-10.
12. Dhalwani NN, Zaccardi F, O'Donovan G, Carter P, Hamer M, Yates T, et al. Association between lifestyle factors and the incidence of multimorbidity in an older English population. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences. 2017;72(4):528-34.
13. Filozof C, Fernandez Pinilla M, Fernández-Cruz A. Smoking cessation and weight gain. Obesity reviews. 2004;5(2):95-103.
14. Schneider S, Huy C, Schuessler M, Diehl K, Schwarz S. Optimising lifestyle interventions: identification of health behaviour patterns by cluster analysis in a German 50+ survey. European Journal of Public Health. 2009;19(3):271-7.
15. Griffin B, Sherman KA, Jones M, Bayl-Smith P. The clustering of health behaviours in older Australians and its association with physical and psychological status, and sociodemographic indicators. Annals of Behavioral Medicine. 2014;48(2):205-14.
16. Liao J, Mawditt C, Scholes S, Lu W, Umeda M, Muniz Terrera G, et al. Similarities and differences in health-related behavior clustering among older adults in eastern and western

countries: A latent class analysis of global aging cohorts. Geriatrics & gerontology international. 2019;19(9):930-7.
17. Hsu H-C, Luh D-L, Chang W-C, Pan L-Y. Joint trajectories of multiple health-related behaviors among the elderly. International journal of public health. 2013;58(1):109-20.
18. Feng Z, van den Broek T, Perra O, Cramm JM, Nieboer AP. Longitudinal health behaviour patterns among adults aged≥ 50 years in China and their associations with trajectories of depressive symptoms. Aging & Mental Health. 2022:1-10.
19. Wister AV, Levasseur M, Griffith LE, Fyffe I. Estimating multiple morbidity disease burden among older persons: a convergent construct validity study to discriminate among six chronic illness measures, CCHS 2008/09. BMC geriatrics. 2015;15:1-12.
20. Skou ST, Mair FS, Fortin M, Guthrie B, Nunes BP, Miranda JJ, et al. Multimorbidity. Nature Reviews Disease Primers. 2022;8(1):48.
21. Harrison C, Britt H, Miller G, Henderson J. Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. BMJ open. 2014;4(7):e004694.
22. Singer L, Green M, Rowe F, Ben-Shlomo Y, Kulu H, Morrissey K. Trends in multimorbidity, complex multimorbidity and multiple functional limitations in the ageing population of England, 2002–2015. Journal of comorbidity. 2019;9:2235042X19872030.
23. Skivington K, Katikireddi S, Leyland A, Hunt K, Mercer S. Risk factors for multimorbidity: A multilevel analysis of a longitudinal cohort from Scotland: Kathryn Skivington. The European Journal of Public Health. 2015;25(suppl_3):ckv167. 020.
24. Northwood M, Ploeg J, Markle-Reid M, Sherifali D. Integrative review of the social determinants of health in older adults with multimorbidity. Journal of Advanced Nursing. 2018;74(1):45-60.
25. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public health reports. 2014;129(1_suppl2):19-31.
26. Fuchs VR. Social determinants of health: caveats and nuances. Jama. 2017;317(1):25-6.
27. Banks J, Phelps, A., Oskala, A., Steptoe, A., Blake, M., Oldfield, Z., Marmot, M., Clemens, S., Rogers, N., Nazroo, J. English Longitudinal Study of Ageing: Waves 0-9, 1998-2019. 2021;([data collection]. 36th Edition.UK Data Service. SN: 5050). doi: http://doi.org/10.5255/UKDA-SN-5050-23.
28. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet. 2007;370(9596):1453-7.
29. Schafer JL. Multiple imputation: a primer. Statistical methods in medical research. 1999;8(1):3-15.
30. Hackett RA, Moore C, Steptoe A, Lassale C. Health behaviour changes after type 2 diabetes diagnosis: Findings from the English Longitudinal Study of Ageing. Scientific reports. 2018;8(1):1-8.
31. Kojima G, Iliffe S, Jivraj S, Walters K. Fruit and Vegetable Consumption and Incident Prefrailty and Frailty in Community-Dwelling Older People: The English Longitudinal Study of Ageing. Nutrients. 2020;12(12):3882.
32. Meier PS, Holmes J, Brennan A, Angus C. Alcohol policy and gender: a modelling study estimating gender-specific effects of alcohol pricing policies. Addiction. 2021;116(9):2372-84.
33. Health Do. Alcohol Guidelines Review – Report from the Guidelines Development Group to the UK Chief Medical Officer. . 2016.
34. Dhalwani NN, O'Donovan G, Zaccardi F, Hamer M, Yates T, Davies M, et al. Long terms trends of multimorbidity and association with physical activity in older English

population. International Journal of Behavioral Nutrition and Physical Activity. 2016;13(1):1-9.
35. Hamer M, de Oliveira C, Demakakos P. Non-exercise physical activity and survival: English longitudinal study of ageing. American journal of preventive medicine. 2014;47(4):452-60.
36. Elias P, McKnight A. Earnings, unemployment and the NS-SEC. A Researcher's Guide to the National Statistics Socio-economic Classification London: Sage. 2003:151-72.
37. Fraser SD, Roderick PJ, Aitken G, Roth M, Mindell JS, Moon G, et al. Chronic kidney disease, albuminuria and socioeconomic status in the Health Surveys for England 2009 and 2010. Journal of Public Health. 2014;36(4):577-86.
38. Tsimpida D, Kontopantelis E, Ashcroft DM, Panagioti M. The dynamic relationship between hearing loss, quality of life, socioeconomic position and depression and the impact of hearing aids: answers from the English Longitudinal Study of Ageing (ELSA). Social psychiatry and psychiatric epidemiology. 2022;57(2):353-62.
39. Mercer S, Furler J, Moffat K, Fischbacher-Smith D, Sanci L. Multimorbidity: technical series on safer primary care: World Health Organization; 2016.
40. DiSantostefano J. International classification of diseases 10th revision (ICD-10). The Journal for Nurse Practitioners. 2009;5(1):56-7.
41. Muthén B, Muthén L. Mplus. Handbook of item response theory: Chapman and Hall/CRC; 2017. p. 507-18.
42. Nylund-Gibson K, Choi AY. Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science. 2018;4(4):440.
43. Chen Q, Luo W, Palardy GJ, Glaman R, McEnturff A. The efficacy of common fit indices for enumerating classes in growth mixture models when nested data structure is ignored: A Monte Carlo study. Sage Open. 2017;7(1):2158244017700459.
44. Tinner L, Wright C, Heron J, Caldwell D, Campbell R, Hickman M. Is adolescent multiple risk behaviour associated with reduced socioeconomic status in young adulthood and do those with low socioeconomic backgrounds experience greater negative impact? Findings from two UK birth cohort studies. BMC public health. 2021;21:1-13.
45. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. Annals of global health. 2019;85(1).
46. Myint PK, Smith RD, Luben RN, Surtees PG, Wainwright NW, Wareham NJ, et al. Lifestyle behaviours and quality-adjusted life years in middle and older age. Age and ageing. 2011;40(5):589-95.
47. Salman A, Sellami M. Do older adults with multimorbidity meet the recommended levels of physical activity? An analysis of Scottish Health Survey. International Journal of Environmental Research and Public Health. 2019;16(19):3748.
48. Wister A, Li L, Whitmore C, Ferris J, Klasa K, Linkov I. Multimorbidity resilience and health behaviors among older adults: A longitudinal study using the Canadian Longitudinal Study on Aging. Frontiers in Public Health. 2022;10:896312.
49. Suhag A. alisha392/Health_Behaviour_Clusters: Version 1.0.0. Zenodo. 2023. doi: https://zenodo.org/record/8157562.
50. Wilson K, Senay I, Durantini M, Sánchez F, Hennessy M, Spring B, et al. When it comes to lifestyle recommendations, more is sometimes less: a meta-analysis of theoretical assumptions underlying the effectiveness of interventions promoting multiple behavior domain change. Psychological bulletin. 2015;141(2):474.
51. Sisti LG, Dajko M, Campanella P, Shkurti E, Ricciardi W, De Waure C. The effect of multifactorial lifestyle interventions on cardiovascular risk factors: a systematic review and meta-analysis of trials conducted in the general population and high risk groups. Preventive Medicine. 2018;109:82-97.

52. Bricca A, Jäger M, Johnston M, Zangger G, Harris LK, Midtgaard J, et al. Effect of in-person delivered behavioural interventions in people with multimorbidity: systematic review and meta-analysis. International Journal of Behavioral Medicine. 2023;30(2):167-89.

Supplementary Material

Table of contents

Section 1: Survey on health behaviour measures

For the survey, researchers who had published at least one relevant article, as the first author, in the last three years on any of the four SNAP behaviours were identified through the research team's personal networks. Out of an opportunistic sample of 20 researchers, 15 experts participated in our survey. The results of the survey indicated a consensus (defined as agreement of 70% or greater) on the most appropriate measure of fruit and vegetable intake, smoking, and physical activity behaviours (for details, see Supplementary Table 1). Although the results did not indicate a clear consensus for alcohol consumption, the choice with the maximum respondent support was discussed between the authors and chosen.

Table S1. Results of the expert survey on health behaviour measures

SMOKING

1. a) Based on the following measures, which do you think would be the most appropriate measure of smoking for use in our study on behavioural clustering?

Ever smoking status (whether or not the person has ever smoked)	0%
Current smoking status (whether or not the person smokes nowadays)	73%
Number of cigarettes smoked per day (roll-ups can be converted to cigarettes)	27%

1. b) Please explain your answer

- The percentage of current smokers is relatively low, which means that to subdivide smokers into amount smoked would further reduce the statistical power to reliably identify clusters associated with smoking.
- If I can only choose one of the above measures, I think current status is most important to look at clustering of behaviour
- Ever smoking could reflect smoking behaviour from decades previous. Number of cigarettes is important, but to me less important than current status. We know that any amount of smoking is harmful, so it matters less that someone is smoking 10 vs 20 and more that they are smoking at all. I also wouldn't know how to advise on what a numerical cut off should be
- Difference between 0 and some smoking is qualitatively different, and won't be captured by looking at CPD among everyone. Current smoking much more important to current clustering than previous smoking (especially a weak measure of ever rather than ever regular - many have smoked a small number without doing so regularly).
- Most of the literature/policy tends to focus on this binary outcome
- Given the dwindling number of smokers in the population, I think that this is all the information that you need. Cigs per day is a good marker of dependence severity though
- Tricky one to start with and largely depends on the research question! I would say current smoking status, closely followed by ever smoked. Easiest for respondents to answer consistently/accurately and most straightforward to interpret

1. c) How would you categorise the number of cigarettes smoked? Into how many categories would you divide the number of cigarettes smoked?

- I would avoid categorising the numbers and would use the number reported as a continuous variable.
- I might suggest having a category for non-smoker (0 cigs) and then basing the next categories 1-5, 6-10, 10-20 and 20 plus
- I think it's best to ask for a number - it is really hard to work with categories in a meaningful way. It should include the option to say none if they don't smoke at all and a figure less than 1 if they do not smoke every day

ALCOHOL CONSUMPTION

The English Longitudinal Study of Ageing collects the following information on alcohol consumption : - Frequency of alcohol consumption over the past 12 months (Almost every day, 5-6 days/week, 3-4 days/week, 1-2 days/week, once or twice a month, once every couple of months, once or twice a year, not at all in the last 12 months) - Frequency of drinking in the past seven days (number of days) - Quantity of beer (pints), wine (glasses) and spirits (ml) consumed over the past seven days.

2. a) Based on the above stated measures, which do you think would be the most appropriate measure for alcohol consumption for use in our clustering analysis? You may even choose a combination of these e.g., a quantity- frequency measure.

Frequency of alcohol consumption over the past 12 months	15%
Frequency of drinking in the past seven days	15%
Quantity of beer (pints), wine (glasses) and spirits (ml) consumed over the past seven days, converted into UK units	38%
None of these (please indicate your answer in the space below)	31%

If you answered 'none of these', please indicate your answer below.

- Both frequency and quantity (it would not allow me to select both). If there is time, I'd suggest breaking the week down into the weekend and week days. If further time is available I'd suggest asking questions about each day to prompt recall.
- I would also ask if their alcohol consumption had changed in the last 7 days compared to their usual alcohol consumption (or is it similar). Quantity-frequency measure - how many units and how often (days per week)
- I would use a quantity-frequency measure averaged over the previous 6 months
- The AUDIT-C would be better as it is a quantity frequency measure that is widely used

2. b) Please explain your answer in 2 a).

- 12 months is more likely to be difficult for the participant to recall and likely to result in more reporting errors. Weekly intake might improve reporting accuracy.
- alc consumption over the past 12 months is the most representative and stable measure. Drinking in the past 7 days and beverage preferences are auxilliary traits that might lead to more detailed clusters being identified, but including these in the first instance risk obscuring the main patterns in clustering that you would want to drive your conclusions
- People are bad at reporting quantity and past 7 day frequency likely to be influenced by seasonal fluctuations.
- Frequency of drinking alone is insufficient to predict risk (someone binge-drinking on one day a week might be at greater risk than someone drinking 1 UK every day)
- This suggestion is based on the literature suggesting that a short recall period is preferable to a long one, particularly when asking participants about routine or frequent events
- Quantity in last 7 days will reflect frequency and will be much better recalled than frequency in last 12 months...
- I've chosen this for pragmatic reasons - it's in the survey. Ideally you'd want to know average normal consumption in units. The past 7 days is fairly reliable though you may have to adjust for seasonality etc. (people drink more at Christmas)
- see above
- Quantity over a period of time takes priority for me over frequency. My justification for this is that we have clear guidelines over recommended units. Most relevant to behavioural clustering (amount/units relevant to risk of harmful drinking)

2. c) Given the measure of alcohol consumption you chose in 2. a), how many categories would you choose?

- I would avoid categories (use continuously), or categorise based on the data collected (e.g. moderation analysis or hierarchial cluster analysis) abst, lower, med, high risk as usual by SARG
- Maybe 5 or so - but this may seem excessive compared to simple smoking categorisation 3
- 3: Lower risk; increasing risk; high risk

2. d) What cut-off points would you use to categorise alcohol consumption?

- 0, 1-14, 14-35 women, 14-50 men, >35 women, >50 men (UK standard units / week)
- 0, 1-14, 14-28, 28+ units or similar
- Daily / almost daily (including 5-6 per week), another category which basically incorporates the old CMO guidance on spreading drinking over 3 or more days, and then a final category which is basically abstainers last two categories)
- Cut-offs used in HSE: 0-13: lower risk; 14-35(women), 14-50(men): increasing risk; 36+(women) 51+(men): high risk number of drinks consumed i.e. 4 pints of beer (I'd then do the conversion to units yourself. Folk don't work in units)

FRUIT AND VEGETABLE INTAKE

The English Longitudinal Study of Ageing study collects information on fruit and vegetable intake on a typical day.

3. a) How would you categorise fruit and vegetable intake and what cut-offs would you use? (Note : Since ELSA has only one measure of nutrition, we are only interested in how this measure of nutrition can be categorised.) The WHO recommends consuming five portions of fruit and vegetables a day. Thus, one way to categorise participants' fruit and vegetable intake would be to divide it into two categories with five portions as the cut-off. i) Do you think it is appropriate to categorise participants into those who meet versus those who do not meet the WHO recommendation?

Yes	77%
No	23%

3. b) Based on the measures used in the ELSA study, how many categories would you divide participants' fruit and vegetable intake into?

- Similar to before, I would avoid categories. But if you need to have them, I would recommend having more than two to capture low and high f/v intake
- Intuitively meeting WHO recommendations makes perfect sense, my only concern is that the average intake of F&V is 2.6 portions in the UK and so you may find it doesn't differentiate groups. Perhaps consider a split based on the median in the population
- 0-1; 2-4; 5+
-
- 3. c) What cut-offs points would you use for each category?
-
- Similar to before, I would avoid categories. But if you need to have them, I would recommend having more than two to capture low and high f/v intake
- Intuitively meeting WHO recommendations makes perfect sense, my only concern is that the average intake of F&V is 2.6 portions in the UK and so you may find it doesn't differentiate groups. Perhaps consider a split based on the median in the population
- 0-1; 2-4; 5+

PHYSICAL ACTIVITY

For physical activity, the English Longitudinal Study of Ageing study collects data on how often (i.e. more than once a week, 2 once a week, 3 one to three times a month, 4 hardly ever, or never?) an individual engages in physical activity at the following intensities :

- **mild (laundry and home repairs)**
- **moderate (gardening, cleaning the car, walking at moderate pace, dancing)**
- **vigorous (e.g., running/ jogging, swimming, cycling, aerobics/gym workout, tennis, and digging with a spade)**

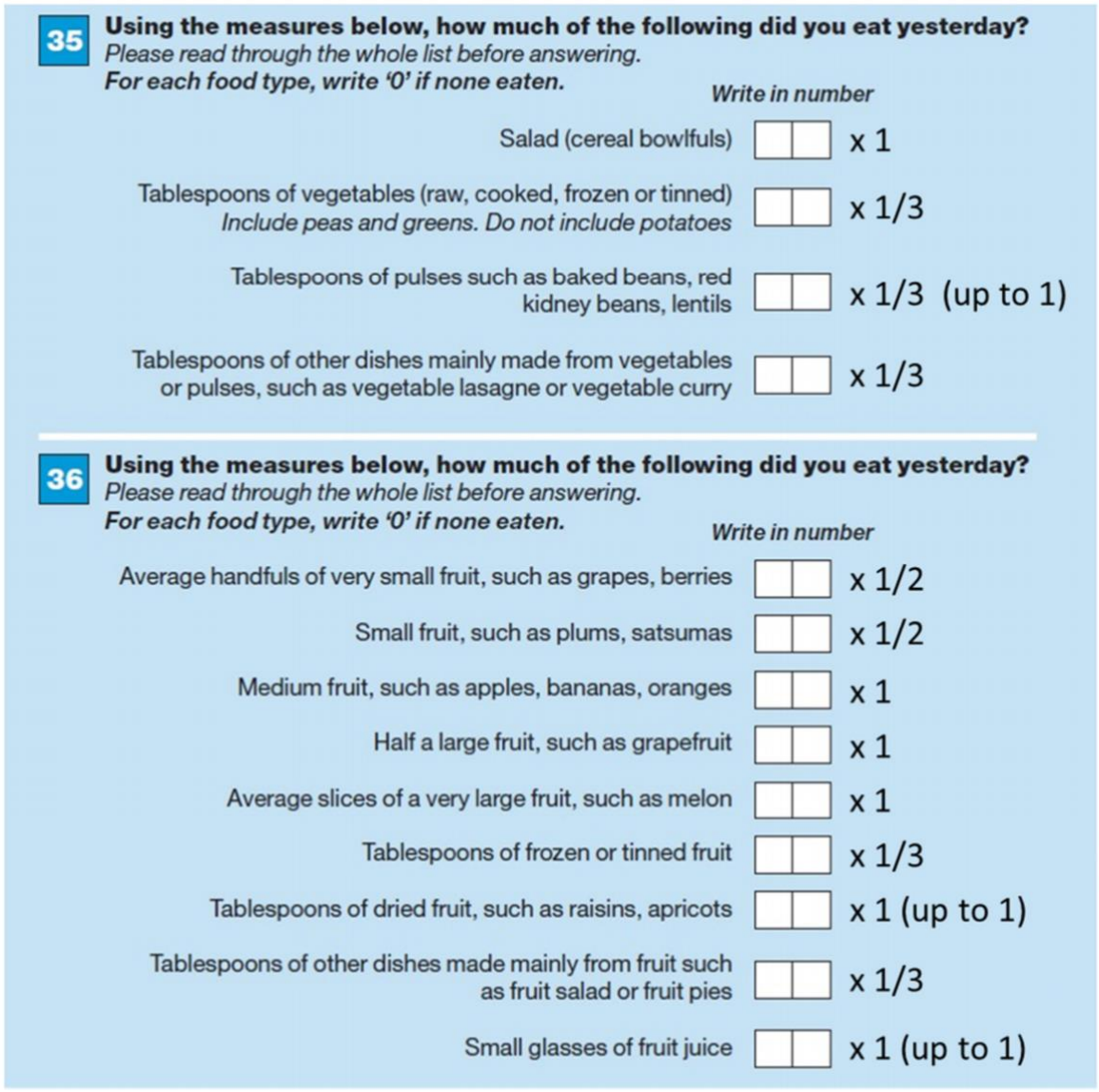
4. a) Based on the above stated measures, some studies have categorised physical activity as : - Sedentary: light exercise 1–3 times a month, no moderate or vigorous activity - Low: light exercise at least once a week but no vigorous activity - Moderate: moderate activity more than once a week, or vigorous activity between once a week to 1–3 times a month - High: vigorous activity more than once a week. i) Do you think this would be an appropriate way to categorise physical activity based on the measures used in the ELSA study? (Note : Since ELSA has only one measure of physical activity, we are only interested in how this measure of physical activity can be categorised.)

Yes	73%
No	27%

4. b) How would you categorise physical activity based on the above stated measures used in the ELSA study?

- I would avoid categories, but if needed, would set moderate as meeting the UK recommended PA guidelines. High would be those reporting physical activity levels above the recommendations.
- The evidence on physical activity in the latest guidance recognises the role that higher volumes of light intensity physical activity can play in health outcomes. This is in addition to moderate and vigorous physical activity. The categories suggested above prioritise vigorous activity and yet a person could engage in one bout of vigorous physical activity per week and be sedentary for the rest of the time. This would not relate to a category of 'highly active'. To overcome this, you could use a combination of mild, moderate and vigorous activity in your categorisation by applying a score to each of the frequencies (1-5) in each of the categories and then using these scores to create categories. In this instance the maximum score would be 15 and the minimum would be 0. Categories could then be 1) A total score of less then 3 = sedentary 2) A total score of 3 - 8 = low 3) A total score of 8 -12 is moderate and 12 plus = high.

Fig S1. Conversion rates for fruit and vegetable intake in Wave 4



35 **Using the measures below, how much of the following did you eat yesterday?**
Please read through the whole list before answering.
For each food type, write '0' if none eaten.

	Write in number	
Salad (cereal bowlfuls)	☐☐	x 1
Tablespoons of vegetables (raw, cooked, frozen or tinned) *Include peas and greens. Do not include potatoes*	☐☐	x 1/3
Tablespoons of pulses such as baked beans, red kidney beans, lentils	☐☐	x 1/3 (up to 1)
Tablespoons of other dishes mainly made from vegetables or pulses, such as vegetable lasagne or vegetable curry	☐☐	x 1/3

36 **Using the measures below, how much of the following did you eat yesterday?**
Please read through the whole list before answering.
For each food type, write '0' if none eaten.

	Write in number	
Average handfuls of very small fruit, such as grapes, berries	☐☐	x 1/2
Small fruit, such as plums, satsumas	☐☐	x 1/2
Medium fruit, such as apples, bananas, oranges	☐☐	x 1
Half a large fruit, such as grapefruit	☐☐	x 1
Average slices of a very large fruit, such as melon	☐☐	x 1
Tablespoons of frozen or tinned fruit	☐☐	x 1/3
Tablespoons of dried fruit, such as raisins, apricots	☐☐	x 1 (up to 1)
Tablespoons of other dishes made mainly from fruit such as fruit salad or fruit pies	☐☐	x 1/3
Small glasses of fruit juice	☐☐	x 1 (up to 1)

Note. "The amount of consumed fruit and vegetables were converted into portions (1 portion = 80 g for both vegetables and fruit) in accordance with the Welsh Health Survey methodology and the "5 A Day" campaign portion size from the National Health Service (NHS)."(Kojima et al., 2020)

Section 2: Supplementary Missing Data Analysis

As seen in Table S2, the included sample had a slightly higher average age (62.9 years vs 60.4, $p<0.001$), a higher proportion of individuals in intermediate level jobs (26.2% vs. 21.6%, $p=0.006$) and professional/managerial occupations (39.3% vs. 30.1%, $p<0.001$). The included sample also had a higher proportion of individuals educated up to degree level or higher (20.9% vs 16.7%, $p=0.015$) and a higher proportion of individuals with parents in intermediate occupations (32.1% vs. 25.8%, $p=0.006$) than the excluded sample. There were no significant differences in sex, other socio-demographic variables, or disease status between the included and excluded samples, except for complex multimorbidity, respiratory disorders and endocrine disorders (see Table S2). In sum, the absolute differences between the included and excluded samples were not substantial, though they achieved significance in some cases due to the relatively large number of participants.

Table S2. Comparison of the included sample (i.e., with no missing data on any sociodemographic variables, n=4759) and the excluded sample (i.e., with missing on at least one sociodemographic variable, n=670)

Baseline socio-demographic characteristics	Included sample (n=4759) N (%)	Excluded sample (n=670) N (%)	Test estimate	p-value
Male	2081 (43.7)	271 (40.4)	2.44	0.118
Female	2678 (56.3)	399 (59.6)	2.44	0.118
Average Age (s.d.)	62.9 (8.1)	60.4 (8.2)	**7.65***	**<0.001***
Parental Occupation – Semi-routine and routine	1299 (27.3)	163 (34.8)	**11.42***	**<0.001***
Parental Occupation – Intermediate	1530 (32.1)	121 (25.8)	**7.68***	**0.006***
Parental Occupation – Professional/managerial	1930 (40.6)	185 (39.4)	0.17	0.676
Occupation – Semi-routine and routine	1646 (34.6)	221 (48.3)	**33.36***	**<0.001***
Occupation – Intermediate	1245 (26.2)	99 (21.6)	**4.28***	**0.039***
Occupation – Professional/managerial	1868 (39.3)	138 (30.1)	**14.30***	**<0.001***
Education – No qualifications	958 (20.1)	172 (26.8)	**14.77***	**<0.001***
Education – Intermediate	2807 (59)	363 (56.5)	1.29	0.256
Education – Degree/higher	994 (20.9)	107 (16.7)	**5.95***	**0.015***
Wealth – First Tertile	1260 (26.5)	127 (33.1)	**7.70***	**0.006***
Wealth – Second Tertile	1615 (33.9)	119 (31.1)	1.19	0.278
Wealth – Third Tertile	1884 (39.6)	137 (35.8)	2.01	0.156
Disease status at Wave 4				
Multimorbidity	1802 (37.9)	263 (39.3)	0.42	0.515
Complex multimorbidity	588 (12.4)	113 (16.9)	**10.23***	**0.001***
Respiratory disorders	548 (11.5)	102 (15.2)	**7.38***	**0.006***
Eye disorders	729 (15.3)	93 (13.9)	0.82	0.365
Musculoskeletal and connective system disorders	1599 (33.6)	222 (33.2)	0.03	0.874
Neoplasms	140 (2.9)	18 (2.7)	0.06	0.814
Circulatory disorders	1870 (39.3)	259 (38.7)	0.06	0.806
Endocrine nutritional & metabolic disorders	334 (7)	68 (10.2)	**7.99**	**0.005***
Disease status at Wave 9				
Multimorbidity	2755 (57.9)	388 (57.9)	0	1.00
Complex multimorbidity	1278 (26.9)	208 (31)	**4.98***	**0.026***
Respiratory disorders	664 (14)	115 (17.2)	**4.80***	**0.028***
Eye disorders	1879 (39.5)	253 (37.9)	0.60	0.441
Musculoskeletal and connective system disorders	2261 (47.5)	323 (48.4)	0.14	0.713
Neoplasms	235 (4.9)	27 (4)	0.84	0.359
Circulatory disorders	2394 (50.3)	350 (52.3)	0.87	0.351
Endocrine nutritional & metabolic disorders	621 (13)	117 (17.5)	**9.51***	**0.002***

Note. The differences for all variables except age were calculated using the two-sample test for equality of proportions. For age, the two-sample Welsh t-test was used to evaluate the differences. ***p<0.05**

Section 3: Supplementary Statistical Analysis

For the likelihood-based tests such as the VLMR-LMR test, a p-value < 0.05 indicates that the model fit has not significantly improved compared to the model with one less class. Lower values of the BIC, CAIC, saBIC, and AWE indicate a better fitting model. However, when assessing the best model fit, particularly in large datasets with multiple indicators, additional classes can often lead to a decrease in the information criterion (ICs) — favouring the more complex model — until no further class can be added due to convergence issues.(Nylund-Gibson and Choi, 2018) Thus, there exists no global minimum. In such cases, the recommendation is to plot the ICs to seek a point of inflection or plateauing.(Nylund-Gibson and Choi, 2018) Finally, entropy and the smallest average latent class posterior probability were also used to quantify how well the model separated individuals into distinct latent classes, with values closer to 1 indicating distinct separation between classes.

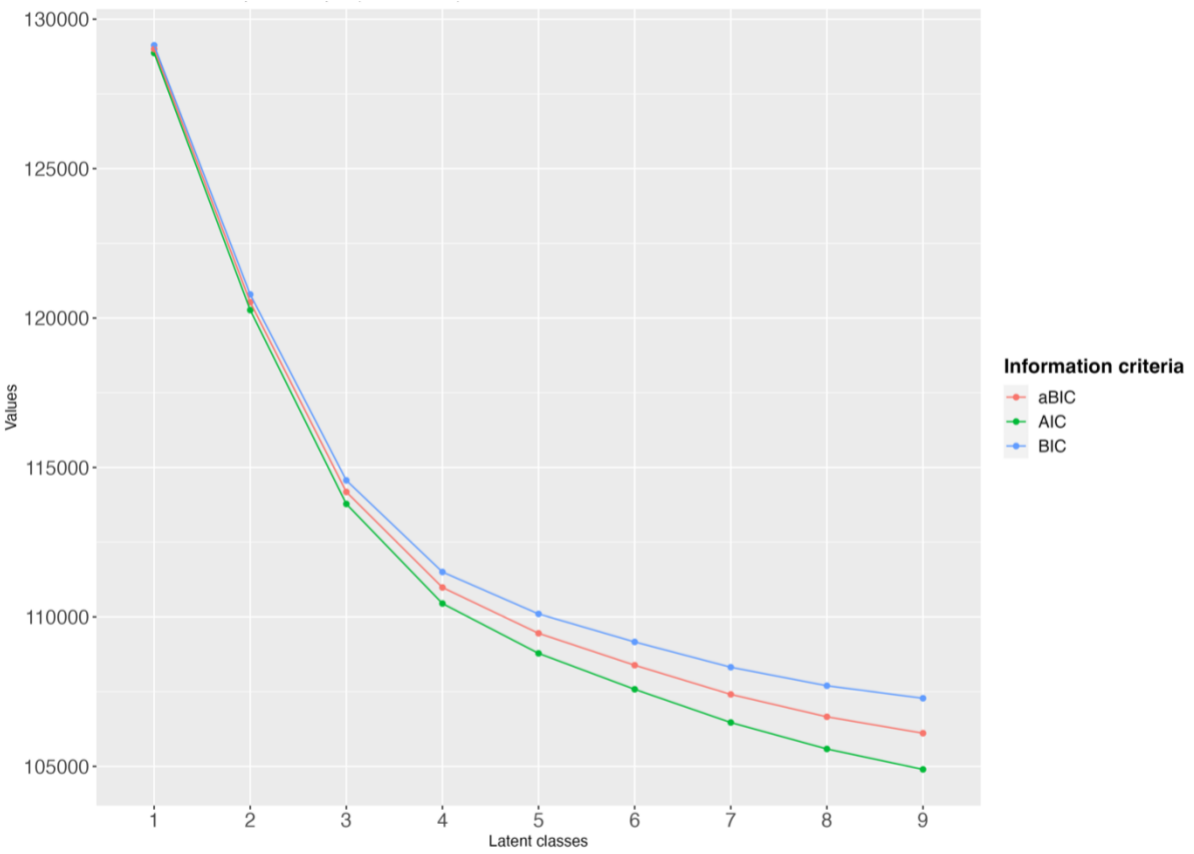
Table S3. Disease status of the participants (N= 4759) at Wave 9 (2018-2019)

Disease status at Wave 9 (2018-2019)		
Multimorbidity, N(%)	Has condition	2755 (57.9)
	Missing	..
Complex multimorbidity, N(%)	Has condition	1278 (26.9)
	Missing	..
Respiratory disorders, N(%)	Has condition	664 (14)
	Missing	1 (<0.1)
Eye disorders, N(%)	Has condition	1879 (39.5)
	Missing	4 (0.1)
Musculoskeletal and connective system disorders, N(%)	Has condition	2261 (47.5)
	Missing	..
Neoplasms, N(%)	Has condition	235 (4.9)
	Missing	1 (0)
Circulatory disorders, N(%)	Has condition	2394 (50.3)
	Missing	..
Endocrine nutritional & metabolic disorders, N(%)	Has condition	621 (13)
	Missing	4 (0.1)
Nervous system disorders, N(%)	Has condition	168 (3.5)
	Missing	4530 (95.2) *

Mental & behavioural issues, N(%)	Has condition	63 (1.3)
	Missing	4688 (98.5) *

* Due to a large proportion of missing data on disease status for two health conditions i.e., nervous system disorders (95.2%), and mental and behavioural issues (98.5%), they were excluded from subsequent analyses.

Fig S2. Elbow plot of Bayesian Information Criteria and other indices for model fitting



Note. BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion

Section 4: Split-half replication

To ensure the reliability of our class solution, the sample (n=4759) was randomly split into halves and separate, sequential latent class models were conducted on both random subsamples. These subsamples were named Sample A (n=2379) and Sample B (n=2380). We then compared the model fit statistics (such as the Bayesian information criterion, sample size adjusted BIC, consistent Akaike information criterion, and loglikelihoods) of the three samples using screeplots. If the final latent class solution (i.e., a seven-class model) was consistently identified as the best fit in all three samples, then this would provide evidence for the replicability of the latent classes identified in the analysis. Fig S3 shows the screeplots for the model fit statistics in the three samples. Screeplots of model fit statistics for one to nine class solutions in the random subsamples and those for the whole sample are shown in Fig S3. In all three samples, the decline in the information criteria levelled off around seven classes, providing initial evidence for replication of the results across the three samples. Additionally, we compared the structures (i.e., the prevalence of latent classes and the probabilities of item responses) of the seven-class model with the five-class and six-class models in the three samples (see Supplementary Figures 4-6). We found that the seven-class model replicated with maximum stability (i.e., had a similar latent class structure) across all three samples. On the other hand, the five-class and six-models were unstable as they did not replicate across all the three samples.

Fig S3. Screeplots of Bayesian information criterion (BIC), Sample size adjusted BIC (SABIC); Consistent Akaike information criterion (CAIC) and loglikelihoods for i) Sample A (n=2379, 50% random split sample), ii) Sample B (n=2380, 50% random split sample) and iii) Complete Sample (n=4759).

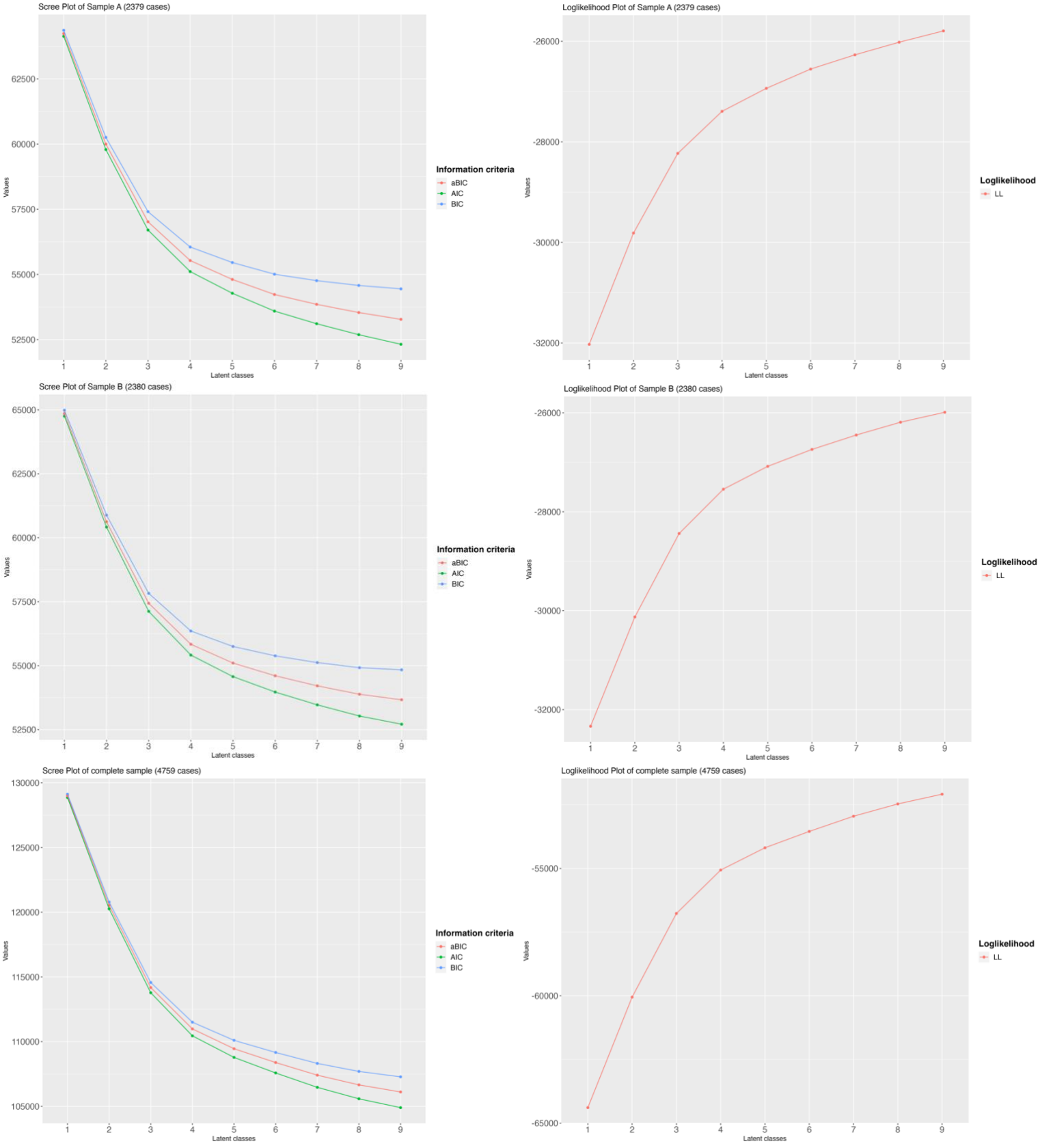
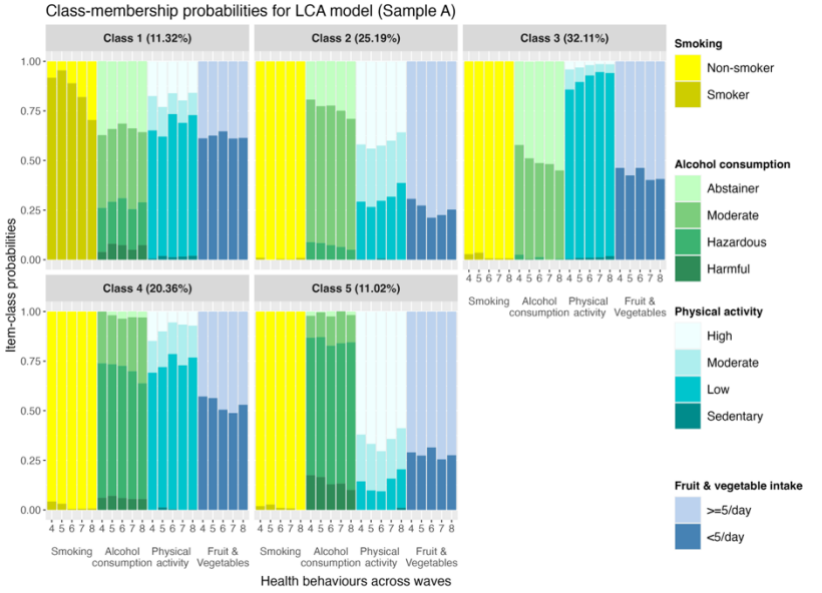
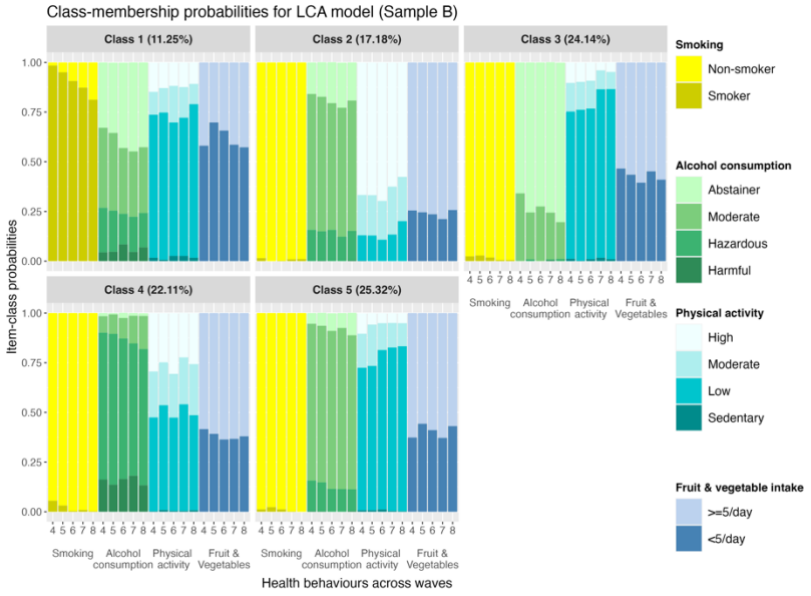


Fig S4. 5-class model solutions for Sample A (n=2379, 50% random split sample), Sample B (n=2380, 50% random split sample) and Complete Sample (n=4759)





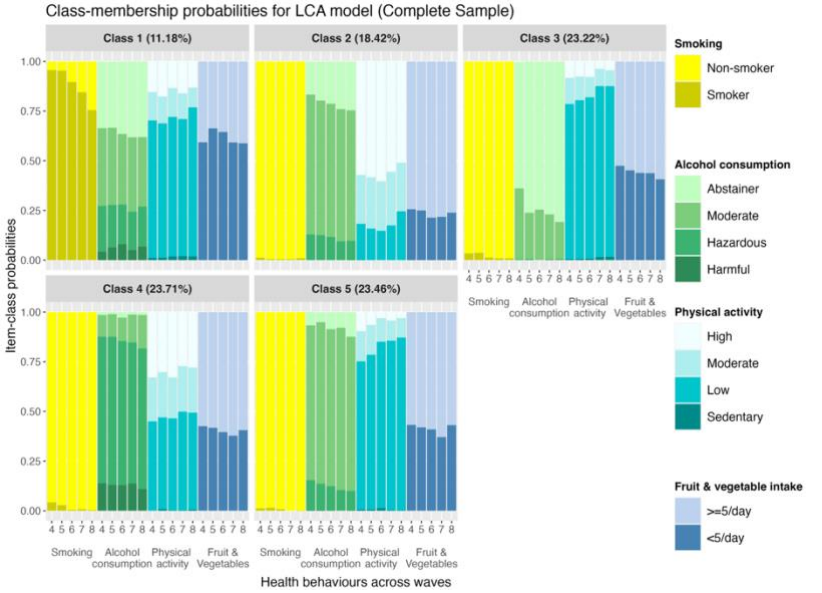
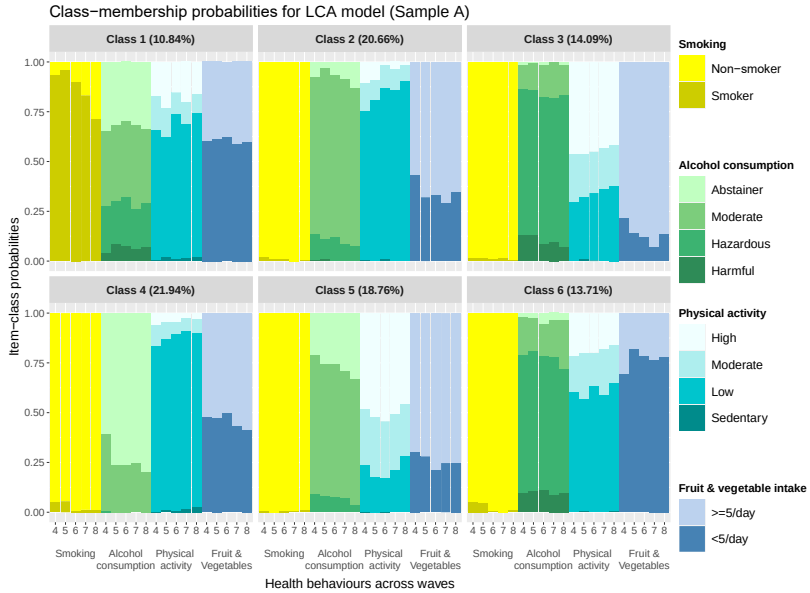
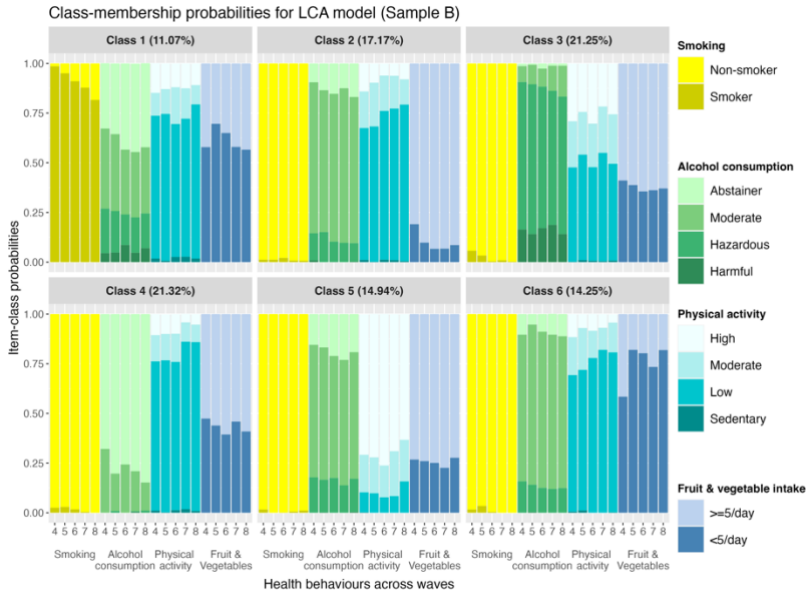


Fig S5. 6-class model solutions for Sample A (n=2379, 50% random split sample), Sample B (n=2380, 50% random split sample) and Complete Sample (n=4759)





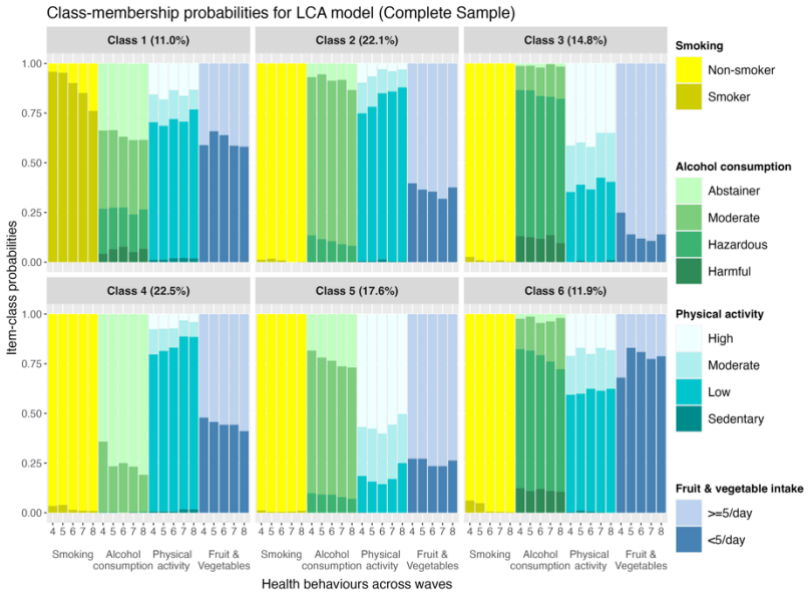


Fig S6. 7-class model solutions for Sample A (n=2379, 50% random split sample), Sample B (n=2380, 50% random split sample) and Complete Sample (n=4759)

Table S5. Pairwise comparison of differences in disease status across latent classes (unadjusted for sociodemographic characteristics and baseline disease status)

Health conditions	Latent classes	Prevalence	1	2	3	4	5	6	7
			Low risk (Ref. class)	**Low risk yet inactive**	**Low risk yet heavy drinkers**	**Abstainers but inactive**	**Poor diet and inactive**	**Inactive, heavy drinkers**	**High-risk smokers**
			(n=13.4 %)	(n=16.8%)	(n=11.4%)	(n=20 %)	(n=12.9 %)	(n=14.5 %)	(n=10.9 %)
Multimorbidity	1	**0.444**		**-0.212**	0.015	**-0.294**	**-0.154**	**-0.15**	**0.129**
				(<0.001)	(0.692)	**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
$\chi2$ (6) = 123.432 , p <0.001									
	2	**0.656**	**0.212**		**0.227**	-0.082	0.058	0.062	-0.083
			(<0.001)		**(<0.001)**	(0.012)	(0.124)	(0.070)	(0.019)
	3	**0.429**	-0.015	**-0.227**		**-0.309**	**-0.169**	**-0.165**	**0.144**
			(0.692)	**(<0.001)**		**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
	4	**0.738**	**0.294**	0.082	**0.309**		**0.14**	**0.144**	**-0.165**
			(<0.001)	(0.012)	**(<0.001)**		**(<0.001)**	**(<0.001)**	**(<0.001)**
	5	**0.598**	**0.154**	-0.058	**0.169**	**-0.14**		0.004	-0.025
			(<0.001)	(0.124)	**(<0.001)**	**(<0.001)**		(0.907)	(0.523)
	6	**0.594**	**0.15**	-0.062	**0.165**	**-0.144**	-0.004		-0.021
			(<0.001)	(0.070)	**(<0.001)**	**(<0.001)**	(0.907)		(0.567)
	7	**0.573**	**0.129**	-0.083	**0.144**	**-0.165**	-0.025	-0.021	
			(<0.001)	(0.019)	**(<0.001)**	**(<0.001)**	(0.523)	(0.567)	
Complex Multimorbidity	1	**0.16**		**-0.174 (<0.001)**	0.051 (0.048)	**-0.269 (<0.001)**	**-0.109 (0.001)**	**-0.099 (<0.001)**	**0.124 (<0.001)**
	2	**0.334**	**0.174**		**0.225**	**-0.095**	0.065	0.075	-0.05
$\chi2$ (6) = 141.267 , p<0.001			**(<0.001)**		**(<0.001)**	**(0.004)**	(0.074)	(0.018)	(0.130)
	3	**0.109**	-0.051	**-0.225**		**-0.32**	**-0.16**	**-0.15**	**0.175**

			(0.048)	**(<0.001)**		**(<0.001)**	**(<0.001)**	**(<0.001)**	**(<0.001)**
	4	**0.429**	**0.269**	**0.095**	**0.32**		**0.16**	**0.17**	**-0.145**
			(<0.001)	**(0.004)**	**(<0.001)**		**(<0.001)**	**(<0.001)**	**(<0.001)**
	5	**0.269**	**0.109**	-0.065	**0.16**	**-0.16**		0.01	0.015
			(0.001)	(0.074)	**(<0.001)**	**(<0.001)**		(0.777)	(0.677)
	6	**0.259**	**0.099**	-0.075	**0.15**	**-0.17**	-0.01		0.025
			(<0.001)	(0.018)	**(<0.001)**	**(<0.001)**	(0.777)		(0.437)
	7	0.284	**0.124**	-0.05	**0.175**	**-0.145**	0.015	0.025	
			(<0.001)	(0.130)	**(<0.001)**	**(<0.001)**	(0.677)	(0.437)	
Respiratory disorders	1	**0.09**		-0.059	-0.003	**-0.097**	-0.016	-0.04	**0.137**
				(0.012)	(0.887)	**(<0.001)**	(0.505)	(0.072)	**(<0.001)**
$\chi2$ (6) = 48.588 , p <0.001									
	2	**0.149**	0.059		0.056	-0.038	0.043	0.019	**0.078**
			(0.012)		(0.015)	(0.131)	(0.104)	(0.406)	**(0.004)**
	3	**0.093**	0.003	-0.056		**-0.094**	-0.013	-0.037	**0.134**
			(0.887)	(0.015)		**(<0.001)**	(0.589)	(0.128)	**(<0.001)**
	4	**0.187**	**0.097**	0.038	**0.094**		**0.081**	0.057	0.04
			(<0.001)	(0.131)	**(<0.001)**		**(0.005)**	(0.018)	(0.164)
	5	**0.106**	0.016	-0.043	0.013	**-0.081**		-0.024	**0.121**
			(0.505)	(0.104)	(0.589)	**(0.005)**		(0.363)	**(<0.001)**
	6	**0.13**	0.04	-0.019	0.037	-0.057	0.024		**0.097**
			(0.072)	(0.406)	(0.128)	(0.018)	(0.363)		**(<0.001)**
	7	**0.227**	**0.137**	**0.078**	**0.134**	0.04	**0.121**	**0.097**	
			(<0.001)	**(0.004)**	**(<0.001)**	(0.164)	**(<0.001)**	**(<0.001)**	
Endocrine, nutritional and	1	**0.074**		**-0.083**	0.027	**-0.173**	**-0.074**	-0.026	**0.065**
				(0.001)	(0.162)	**(<0.001)**	**(0.004)**	(0.181)	**(0.005)**

metabolic	2	**0.157**	**0.083**		**0.11**	**-0.09**	0.009	0.057	-0.018
disorders			**(0.001)**		**(<0.001)**	**(0.001)**	(0.764)	(0.019)	(0.500)
$\chi 2$ (6) = 90.007 , p=<0.001	3	**0.047**	-0.027	**-0.11**		**-0.2**	**-0.101**	-0.053	**0.092**
			(0.162)	**(<0.001)**		**(<0.001)**	**(<0.001)**	(0.014)	**(<0.001)**
	4	**0.247**	**0.173**	**0.09**	**0.2**		**0.099**	**0.147**	**-0.108**
			(<0.001)	**(0.001)**	**(<0.001)**		**(0.003)**	**(<0.001)**	**(<0.001)**
	5	**0.148**	**0.074**	-0.009	**0.101**	**-0.099**		0.048	-0.009
			(0.004)	(0.764)	**(<0.001)**	**(0.003)**		(0.075)	(0.763)
	6	**0.1**	0.026	-0.057	0.053	**-0.147**	-0.048		0.039
			(0.181)	(0.019)	(0.014)	**(<0.001)**	(0.075)		(0.110)
	7	0.139	**0.065**	-0.018	**0.092**	**-0.108**	-0.009	0.039	
			(0.005)	(0.500)	**(<0.001)**	**(<0.001)**	(0.763)	(0.110)	
Circulatory disorders $\chi 2$ (6) = 66.513 , p= <0.001	1	**0.411**		**-0.125**	0.015	**-0.221**	-0.098	**-0.127**	**0.116**
				(0.001)	(0.683)	**(<0.001)**	(0.012)	**(<0.001)**	**(0.002)**
	2	**0.536**	**0.125**		**0.14**	**-0.096**	0.027	-0.002	-0.009
			(0.001)		**(<0.001)**	**(0.005)**	(0.485)	(0.947)	(0.812)
	3	**0.396**	-0.015	**-0.14**		**-0.236**	**-0.113**	**-0.142**	**0.131**
			(0.683)	**(<0.001)**		**(<0.001)**	**(0.004)**	**(<0.001)**	**(<0.001)**
	4	**0.632**	**0.221**	**0.096**	**0.236**		**0.123**	**0.094**	**-0.105**
			(<0.001)	**(0.005)**	**(<0.001)**		**(0.001)**	**(0.005)**	**(0.003)**
	5	**0.509**	0.098	-0.027	**0.113**	**-0.123**		-0.029	0.018
			(0.012)	(0.485)	**(0.004)**	**(0.001)**		(0.450)	(0.644)
	6	**0.538**	**0.127**	0.002	**0.142**	**-0.094**	0.029		-0.011
			(<0.001)	(0.947)	**(<0.001)**	**(0.005)**	(0.450)		(0.765)
	7	0.527	**0.116**	-0.009	**0.131**	**-0.105**	0.018	-0.011	

			(0.002)	(0.812)	**(<0.001)**	**(0.003)**	(0.644)	(0.765)	
Eye disorders	1	**0.323**		**-0.129**	0.012	**-0.163**	**-0.113**	-0.056	-0.01
$\chi 2$ (6) = 57.967 , p= <0.001				**(<0.001)**	(0.722)	**(<0.001)**	**(0.003)**	(0.087)	(0.751)
	2	**0.452**	**0.129**		**0.141**	-0.034	0.016	0.073	**-0.139**
			(<0.001)		**(<0.001)**	(0.326)	(0.677)	(0.034)	**(<0.001)**
	3	**0.311**	-0.012	**-0.141**		**-0.175**	**-0.125**	-0.068	0.002
			(0.722)	**(<0.001)**		**(<0.001)**	**(0.001)**	(0.057)	(0.960)
	4	**0.486**	**0.163**	0.034	**0.175**		0.05	**0.107**	**-0.173**
			(<0.001)	(0.326)	**(<0.001)**		(0.196)	**(0.001)**	**(<0.001)**
	5	**0.436**	**0.113**	-0.016	**0.125**	-0.05		0.057	**-0.123**
			(0.003)	(0.677)	**(0.001)**	(0.196)		(0.135)	**(0.001)**
	6	**0.379**	0.056	-0.073	0.068	**-0.107**	-0.057		-0.066
			(0.087)	(0.034)	(0.057)	**(0.001)**	(0.135)		(0.051)
	7	0.313	-0.01	**-0.139**	0.002	**-0.173**	**-0.123**	-0.066	
			(0.751)	**(<0.001)**	(0.960)	**(<0.001)**	**(0.001)**	(0.051)	
Musculoskeletal disorders	1	**0.392**		-0.152	0.01	-0.191	-0.075	-0.058	0.106
$\chi 2$ (6) = 58.410 , p= <0.001				(0.544)	(0.256)	(0.254)	(0.146)	(0.402)	(0.607)
	2	**0.544**	0.152		0.162	-0.039	0.077	0.094	-0.046
			(0.544)		(0.540)	(0.556)	(0.334)	(0.823)	(0.947)
	3	**0.382**	-0.01	-0.162		-0.201	-0.085	-0.068	0.116
			(0.256)	(0.540)		(0.957)	(0.725)	(0.701)	(0.529)
	4	**0.583**	0.191	0.039	0.201		0.116	0.133	-0.085
			(0.254)	(0.556)	(0.957)		(0.682)	(0.712)	(0.540)
	5	**0.467**	0.075	-0.077	0.085	-0.116		0.017	0.031
			(0.146)	(0.334)	(0.725)	(0.682)		(0.457)	(0.349)

6	**0.45**	0.058	-0.094	0.068	-0.133	-0.017		0.048
		(0.402)	(0.823)	(0.701)	(0.712)	(0.457)		(0.789)
7	0.498	0.106	-0.046	0.116	-0.085	0.031	0.048	
		(0.607)	(0.947)	(0.529)	(0.540)	(0.349)	(0.789)	

Note. The omnibus Wald test results were significant for all eight disease outcomes was significant except Neoplasm $\chi 2$ (6) = 7.263, p= 0.297. Significant omnibus Wald test results indicate that the prevalence of these health outcomes differed across clusters and were followed up with pairwise Wald tests. The estimates shown above are the absolute differences in proportions of participants having a given health outcome in the class (in row) minus the class (in column). **Bold values are statistically significant at the Bonferroni-corrected significance level (p=0.007)** and indicate the two-tailed p-values for pairwise Wald test for differences in proportion for the class (in row) minus the class (in column).

Section 5: STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Abstract, second paragraph	"Longitudinal study"
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract, paragraph 2-3	
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction, paragraphs 1-5	
Objectives	3	State specific objectives, including any prespecified hypotheses	Introduction, paragraph 6	"Our research analyses data from a longitudinal panel of older adults in England to: i) explore how the SNAP behaviours cluster over time in older adults, ii) investigate how membership in different behavioural clusters varies by socio-demographic characteristics, and iii) examine which, if any, behavioural clusters are prospectively associated with multimorbidity over time. "
Methods				
Study design	4	Present key elements of study design early in the paper	Methods, paragraph 1	
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods, paragraph 1	"We analysed secondary data from the English Longitudinal Study of Ageing (ELSA) – a nationally representative, ongoing panel study of community-dwelling adults aged 50 and over at baseline, in England [27]. ELSA collects biennial data on mental and physical health, finances, and

				attitudes around ageing using computer-assisted interviews and questionnaires [27].”
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	Methods, paragraph 2-3	“Our analysis used data from 5,429 respondents to the core questionnaire across six waves from Wave 4 (2008-2009) to Wave 9 (2018-2019). We applied the longitudinal weights that were provided with the dataset and had been derived using information spanning from Wave 4 to Wave 9 to reduce drop-out bias. Wave 4 was selected as the baseline because, although data on health behaviours was available from Wave 3, longitudinal weights were only available from Wave 1 or Wave 4. Choosing Wave 3 as the baseline would have resulted in the loss of data on approximately 2000 participants due to longitudinal weighting. Participants (n=670) with missing values on socio-demographic variables were removed using listwise deletion, leaving a final sample of 4759 participants (87.6% of the original sample; see Fig 1).”
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/a	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Methods, paragraph 5-13	

Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods, paragraph 5-13	
Bias	9	Describe any efforts to address potential sources of bias	Methods, paragraph 4, Supplementary Section 2	"To assess the potential impact of excluding participants with missing data, we compared the included and excluded samples (for details, see Supplementary Section 2). Overall, the absolute differences between the included and excluded samples were not substantial, though differences for some socio-demographic variables (average age, tertiary education, intermediate and professional/managerial occupations) and disease status (complex multimorbidity, respiratory disorders and endocrine disorders) achieved significance due to the relatively large number of participants."
Study size	10	Explain how the study size was arrived at	Methods, paragraph 2-3, Figure 1	"Our analysis used data from 5,429 respondents to the core questionnaire across six waves from Wave 4 (2008-2009) to Wave 9 (2018-2019). We applied longitudinal weights from Wave 4 to reduce drop-out bias. Wave 4 was selected as the baseline because although health behaviour data were available starting in Wave 3, longitudinal weights were only available for Waves 1 and 4. Choosing Wave 3 as the baseline would have resulted in the loss of data on

	approximately 2000 participants due to longitudinal weighting. Participants (n=670) with missing values on one or more socio-demographic variables were removed using listwise deletion, leaving a final effective sample of 4759 participants (87.6% of the original sample)”

Continued on next page

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Methods, paragraph 5-13	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Methods, paragraph 14-17	" RMLCA was used to examine whether there were distinct classes of respondents who had similar patterns of SNAP behaviours over time. RMLCA was chosen as it adopts a probabilistic model-based approach for capturing the number and composition of clusters, handles categorical data well, and allows for reliable interpretation and replication of patterns uncovered in the data. MPlus v8.5 software and R version v4.0.3 [41, 42] was used to conduct the RMLCA. A two-stage approach was used...."
		(*b*) Describe any methods used to examine subgroups and interactions	N/A	
		(*c*) Explain how missing data were addressed	Methods, paragraph 3	" Participants (n=670) with missing values on socio-demographic variables were removed using listwise deletion, leaving a final sample of 4759 participants (87.6% of the original sample; see Fig 1). We chose to use listwise deletion because the MPlus v8.5 software package does not support handling missing data for socio-demographic predictors in a latent class analysis (see Figure 1). More specifically, data was missing for occupation (n=212), education (n=28), wealth (n=287), and parental occupation (n=201). Given that no socio-demographic variable had more than 5% missing data, a threshold below which multiple imputation is

				deemed less beneficial, we favoured a complete case analysis [29]."
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	N/a	
		(*e*) Describe any sensitivity analyses	Methods paragraph 15, Supplementary Section 4	"To assess the reliability of the class solution, we conducted a split-half replication where the sample was randomly split in half, and the above RMLCA was performed separately on these split samples to see if the solution for the full sample was replicated between these smaller splits."
Results				
Participants	13	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	N/a	
		(b) Give reasons for non-participation at each stage	N/a	
		(c) Consider use of a flow diagram	Figure 1	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Results paragraph 1, Table 2	"Participants had an average age of 62.9 years (SD = 8.1) and approximately half were female (56.3%; see Table 2 for baseline demographic data). The sample's engagement in health behaviours across waves is shown in Table 3. The body system disorders with the highest prevalence were: multimorbidity (57.9%), circulatory disorders (50.3%), and disorders of the musculoskeletal and connective system (47.5%) (see Supplementary Table 3)."
		(b) Indicate number of participants with missing data for each variable of interest	Supplementary Table 2	

		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/a
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	Supplementary Table 2
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/a
		Cross-sectional study—Report numbers of outcome events or summary measures	N/a
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 5, Table 6, Supplementary Table 5
		(*b*) Report category boundaries when continuous variables were categorized	N/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/a

Continued on next page

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Supplementary Section 4, Supplementary figures 3-6	
Discussion				
Key results	18	Summarise key results with reference to study objectives	Discussion, paragraph 1-4	
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion, paragraph 6	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion paragraphs 2-4, 8	
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion, paragraph 6	"Finally, missing data are unavoidable in general population cohorts such as ELSA. We excluded participants with missing sociodemographic data at baseline. Those who were included were slightly older, were better educated, and were more likely to have more intermediate and professional level jobs than those who were excluded, meaning selection bias due to non-random exclusion is possible. This may limit the generalisability of our findings. Finally, as there are few ethnic minority participants in ELSA, our findings may not generalise to non-white populations."
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Funding section	

References

1. Kojima G, Iliffe S, Jivraj S, Walters K. Fruit and Vegetable Consumption and Incident Prefrailty and Frailty in Community-Dwelling Older People: The English Longitudinal Study of Ageing. Nutrients. 2020;12(12):3882.

2. Nylund-Gibson K, Choi AY. Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science. 2018;4(4):440.

Chapter 5: Longitudinal clustering of health behaviours and their association with multimorbidity: Evidence from Understanding Society (UKHLS)

Overview: This paper presents a conceptual replication of the study using the ELSA data and aims to examine the same research questions but in the general adult sample in the UK, thereby covering a wider age-range and earlier life stages. Since this study has been written up as an independent manuscript, there could be some repetition of the material stated in the background chapter of this thesis.

Introduction

Multimorbidity, the coexistence of two or more chronic conditions, is a growing global challenge, affecting 42% of the adult population (Ho et al., 2022a). Exacerbated by an ageing population and declining mortality rates, multimorbidity in the United Kingdom (UK) is now occurring earlier in the life course with a higher prevalence in recent cohorts (Head et al., 2020; Kingston, Comas-Herrera and Jagger, 2018; Divo, Martinez and Mannino, 2014). While age remains the primary driver of multimorbidity, more individuals under the age of 65 are affected by multimorbidity in absolute numbers compared to those aged 65 and above, largely due to this age group making up a larger proportion of the population (Skou et al., 2022). Thus, it is evident that multimorbidity is not just a feature of ageing, highlighting an urgent need for prevention strategies to mitigate its impact on both individuals and healthcare systems.

Lifestyle factors such as smoking, unhealthy nutrition, alcohol consumption, and physical inactivity (collectively known as SNAP behaviours) play a contributing role to multimorbidity across age groups. However, assessing the importance of these factors is challenging due to heterogeneity in the literature. For instance, a Canadian study examining the association between SNAP behaviours and multimorbidity, found smoking to be the most important factor but also reported that the presence of combinations of unhealthy factors (e.g., smoking and physical inactivity) increased the risk of multimorbidity (Fortin et al., 2014). This study did not show an increased risk of multimorbidity with physical inactivity, yet others have, such as a study using data from the China Health and Retirement Longitudinal Study, which showed that low levels of physical activity were associated with a 45% increased risk of multimorbidity (He et al., 2021). Similarly, a recent Australian study involving 53,867 participants (45–64 years) from the 45 and Up Study who were free of eleven predefined chronic conditions at baseline (2006–2009) showed that the top multimorbidity predictors were age, body mass index, and chicken and red meat intake, as well as smoking in men, but that other behavioural factors like physical activity, alcohol consumption and sleep duration were also important (Shang et al., 2020). A study from India of 699,686 women showed that women who smoked or chewed tobacco had an 87% higher risk of multimorbidity, while those who consumed alcohol had an 18% greater risk (Mishra, Srivastava and Murthy, 2021).

However, the aforementioned studies have either investigated individual risk behaviours, examined interactions between two behaviours, or used a cumulative risk index by summing up the number of risky behaviours an individual engages in. Investigating single health behaviours is insufficient to provide insights into the distinct ways that multiple health behaviours combine into meaningful lifestyle patterns to influence disease outcomes. Additionally, a cumulative risk index does not tell us which combinations of behaviours entail

the greatest risk. Thus, while examining behaviours individually or using the cumulative risk index may be valuable when it comes to measuring population-attributable risk of disease, such methods leave important gaps in the evidence available for designing preventive interventions (i.e. whom and which behaviours or combinations of behaviour to target; Shang et al., 2020).

Recognising the limitations of assessing lifestyle factors individually or through cumulative risk indices, emerging evidence underscores the importance of understanding how these factors cluster within specific population subgroups (Noble et al., 2015). This recognition is important given the observed disparities in risk behaviours and health impact among different demographic groups (Marmot et al., 2012). For example, males and individuals facing greater social disadvantage often exhibit riskier patterns of behaviour (Noble et al., 2015). Individuals from socioeconomically deprived backgrounds are also disproportionately vulnerable to unhealthy lifestyle factors, often experiencing greater risks of harm from a given exposure compared to those from more affluent backgrounds, especially when deprivation persists across their lifespan (Katikireddi et al., 2017; Freisling et al., 2020). Thus, addressing and preventing multimorbidity requires not only targeting individual behaviours but also considering broader social and economic determinants that can exacerbate the risk of disease.

In this context, latent class analysis (LCA) — a modelling technique that aim to identify homogeneous, mutually exclusive classes (or sub-groups) in a population based on similar patterns of responses — is a valuable tool for identifying subgroups with distinct combinations of multiple risk factors (Lanza, Tan and Bray, 2013). In particular, it has a number of advantages for examining risk behaviours to identify subgroups at high risk of disease. First, since LCA is a latent variable modelling technique, it can allow us to detect unique combinations of health risk behaviours that would not have been observable using standard

single variable analyses or even cumulative indices. By identifying specific latent (i.e. hidden) subgroups of participants based on the combinations of behaviours that they engage in, LCA allows us to examine qualitative differences in behavioural profiles across subgroups (Chaoqun et al., 2022). This, in turn, may present opportunities for identifying groups with riskier patterns of behaviour that can be targeted in disease-prevention programs (Noble et al., 2015). Secondly, using LCA to analyse the associations between behaviour and disease can offer a more comprehensive approach to risk assessment. By representing the risks associated with patterns of multiple health behaviours in a single latent variable, LCA eliminates the need to examine risk behaviours individually or rely on multiple interaction terms in traditional regression models, thus reducing the probability of Type I errors. Thirdly, given the absence of a monotonic relationship among health risk behaviours—meaning increased engagement in one behaviour does not consistently correlate with increased or decreased engagement in another—using LCA may highlight associations that conventional regression models might miss, as they cannot fully account for the complex and sometimes unpredictable interactions among these behaviours (Bolin and Lindgren, 2016). Finally, the exploratory nature of LCA allows for the mapping of behavioural profiles without predetermined assumptions about their co-occurrence or how they may impact health outcomes. Starting with hypotheses about behaviour combinations that exist could inadvertently limit exploration and overlook novel connections. All the advantages mentioned above also extend to the longitudinal extension of LCA, known as repeated measures latent class analysis (RMLCA). Although increasingly used in developmental research and to study patterns of individual risk behaviours, RMLCA remains relatively underutilised in other areas (Karamouzian et al., 2024; Lanza and Cooper, 2016). In light of the vast repositories of population health data and an ageing population characterised by high rates of multimorbidity, unsupervised learning techniques like latent class analysis hold untapped potential for uncovering subgroups at high risk of disease.

In light of the recognised significance of identifying high-risk subgroups using LCA, it is useful to acknowledge that despite these advancements, our understanding of how these risk behaviour clusters are associated with disease outcomes, particularly multimorbidity, remains limited (Leventhal, Huh and Dunton, 2014; Tegegne, Islam and Maddison, 2022). While LCA offers valuable insights for identifying sub-groups in various contexts (Porcu and Giambona, 2017), its application in elucidating the complex relationships between risk behaviours and disease outcomes is still underexplored. This gap in the literature is consequential because some combinations of risk behaviours can have a multiplicative, deleterious effect on health outcomes that goes beyond the additive effect arising from each separate behaviour (Tegegne, Islam and Maddison, 2022). Yet, to date, only one study has examined how clusters of risk behaviours are associated with multimorbidity (Suhag, Webb and Holmes, 2024). The study by Suhag, Webb and Holmes (2024) identified seven clusters: (i) *Low-risk* (13.4%), (ii*) Low-risk yet inactive* (16.8%), (iii) *Low-risk yet heavy drinkers* (11.4%), (iv) *Abstainer yet inactive* (20.0%), (v) *Poor diet and inactive* (12.9%), (vi) *Inactive, heavy drinkers* (14.5%), and (vii) *High-risk smokers* (10.9%). Interestingly, these clusters exhibited fairly stable behaviour patterns over time and had different socio-demographic profiles.

Consistent with alcohol consumption patterns in the UK (Meier et al., 2021), the two clusters of heavy drinkers were predominantly male. Notably, people in the clusters characterised by physical inactivity (in combination with other risky behaviours) had lower levels of education and wealth. Crucially, these clusters also differed in disease prevalence, yielding some expected and unexpected findings. For instance, compared to other clusters, people in the *Low-risk* and *Low-risk yet heavy drinkers* had a lower prevalence of all health conditions studied,

while *High-risk smokers* were most likely to have respiratory disorders, as might be expected (Banks et al., 2019).

However, the investigation yielded a particularly unexpected finding: the *Abstainer but inactive* cluster comprised mostly women and had the highest prevalence of multimorbidity, complex multimorbidity (i.e. the co-occurrence of three or more chronic conditions affecting three or more different body systems within an individual), and endocrine disorders. That the cluster characterised by physical inactivity (but no other risk behaviours) was associated with worse health outcomes than clusters characterised by multiple risk behaviours suggested that the relationship between behaviour and health outcomes is more complex than a linear dose–response relationship. Moreover, the persistence of this counterintuitive association, despite adjusting for key socio-demographic factors (age, sex, education, wealth, occupation), is what makes these findings particularly noteworthy.

In essence, the study highlighted that the number or combination of risk behaviours alone could not explain why some clusters had worse health outcomes than others (Suhag, Webb and Holmes, 2024). Yet, since the study focused on older adults in England, its direct applicability to the general adult population is limited. Exploring a younger population could provide valuable insights into the diverse aetiological pathways and progression of chronic diseases, especially considering potential variations across different life stages (Ben-Shlomo and Kuh, 2002). Given the significance of critical periods of exposure and the accumulation of risks over time, which may occur through 'chains of risk'—where one adverse (or beneficial) exposure or experience leads to another, age plays a key role in the development of multimorbidity (Ben-Shlomo and Kuh, 2002). Another compelling argument for examining how risk behaviours cluster over time speaks to the evidence suggesting unstable risk patterns among young people and the potential for early interventions to yield significant lifetime health benefits (Mahalik et

al., 2013; Daw, Margolis and Wright, 2017; Campbell et al., 2014). Furthermore, conducting a conceptual replication to investigate the association between risk behaviour clusters over time and multimorbidity within the UK's cultural context, albeit covering a broader age-range, as introduced by Suhag, Webb and Holmes (2024), could strengthen the credibility of the findings.

Thus, the present research analyses data from a longitudinal, nationally-representative sample of adults in the UK to: i) explore whether and how the SNAP behaviours cluster over time in adults, ii) investigate whether and how membership in different behavioural clusters varies by socio-demographic characteristics, and iii) examine which, if any, behavioural clusters are prospectively associated with multimorbidity over time.

Methods

Study design

We conducted a secondary analysis of data from the UK Household Longitudinal Study (UKHLS)—also known as Understanding Society— which is a nationally representative panel survey of around 40,000 households followed annually since 2009–2010 (for a detailed description of each subcomponent, see: Fumagalli, Knies and Buck, 2017). UKHLS collects information on multiple topics, including health, education, income and social life. The sample is representative of the UK population, comprising clustered, stratified samples of households in England, Scotland, and Wales and a non-clustered, systematic random sample in Northern Ireland (for details, see: Fumagalli, Knies and Buck, 2017). Further, the UKHLS comprises multiple sample components: the General Population Sample (GPS), the Ethnic Minority Boost Sample (EMBS), the Immigrant and Ethnic Minority Boost Sample (IEMBS), and the British Household Panel Survey sample (BHPS) (for a detailed description of each subcomponent, see: McFall, Nandi and Platt, 2017). To date, eligible participants have been assessed annually through 13 waves of data collection since 2009-2010, where each wave is conducted over a 24-month period with face-to-face and self-completion computer-aided personal interviews.

Ethical approval for the UKHLS data used in this study was obtained by the University of Essex Ethics Committee (Institute for Social and Economic Research and Essex, 2023). The data is fully anonymised and freely available to researchers through the End User Licence online registration (Institute for Social and Economic Research and Essex, 2023).

Sample Selection and Exclusion Criteria

The analysis used data from 18318 respondents aged 16+ years who responded to the main adult questionnaire across all seven waves of data collection between Wave 7 (2015-2017) and Wave 13 (2021-2023). Wave 7 was used as the baseline wave since prior waves used inconsistent measures for three out of the four SNAP risk behaviours we aimed to analyse (i.e. all behaviours except smoking had inconsistent measures).

Participants (n=310) with missing values on socio-demographic variables, i.e. participants with missing data on their age (n=2), education (n=194) and household income (n=99), at the baseline (i.e. Wave 7) were removed using listwise deletion, leaving a final sample of 18008 participants (98.3 % of the original sample). In dealing with missing data, we chose to use listwise deletion because the MPlus v8.5 software package used for the analyses does not support handling missing data on predictors in regression models (see Figure 1).

To produce nationally representative results, all analyses utilised longitudinal survey weights computed by UKHLS to account for participant non-response, potential sampling biases and the unequal probability of being sampled. Further, UKHLS has implemented automatic sample replenishment rules to ensure a representative sequence of new cohorts continuously join the study (Fumagalli, Knies and Buck, 2017). For example, if a survey participant leaves a household, efforts are made to track and include them in their new household in the survey. Additionally, new members who join a sample household, including children, are added to the samples. Although some attrition is unavoidable, the longitudinal weighted sample starting at Wave 7 is likely to be broadly representative of the UK population, with the exception of new immigrants who arrive in the country after the study has commenced.

Health behaviour measures

Data on the four SNAP risk behaviours were taken from Waves 7, 9, 11, and 13. In UKHLS, data on smoking, fruit and vegetable consumption, and physical activity participation were collected by a main CAPI questionnaire (Computer Assisted Personal Interview, administered face-to-face by an interviewer), whereas alcohol consumption was collected by a separate confidential self-completion questionnaire.

To categorise responses to the four risk behaviours, we used government recommendations or contemporary public health advice to classify levels of risky engagement as detailed below.

Smoking

We used current smoking status as a measure of smoking. The data comprised participants' binary (yes/no) responses to the question 'Do you smoke cigarettes?'. This question excluded the consumption of electronic cigarettes.

Fruit and Vegetable intake

We used a binary measure to assess the daily intake of fruits and vegetables. Participants were asked, 'On the days when you eat vegetables, how many portions (i.e. 3 heaped tablespoons) do you eat? Please do not include potatoes.' and 'On the days when you eat fruit, how many portions (e.g., an apple, an orange, some grapes) do you eat?' These responses were aggregated to form a composite variable for daily fruit and vegetable intake for each survey wave, which was subsequently categorised into two groups in line with the recommended UK guidelines: less than 5 portions per day or 5 or more portions per day (Appleton et al., 2018).

Alcohol consumption

Alcohol use was evaluated using the AUDIT-C, which is a three-item version of the Alcohol Use Disorders Identification Test and serves as a screening tool to identify hazardous drinking

or active alcohol use disorders (Bécares, 2021). AUDIT-C scores range from 0 to 12, with scores of 0 indicating no alcohol use. Alcohol frequency was assessed by the question, "Thinking about the past 12 months, how often did you have a drink containing alcohol?" and responses were measured on a five-point scale (0 = Never, 1 = Monthly or less, 2 = 2-4 times per month, 3 = 2-3 times per week, 4 = 4+ times per week). Alcohol quantity was determined by the question, "How many drinks did you have on a typical day when you were drinking?" and responses were measured on a five-point scale (0 = 1-2 drinks, 1 = 3-4 drinks, 2 = 5-6 drinks, 3 = 7-9 drinks, 4 = 10+ drinks). The frequency of high episodic drinking was assessed by the question, "How often did you consume 6 or more units (if female) or 8 or more units (if male) on a single occasion in the past year?" and responses were measured on a five-point scale (0 = Never, 1 = Less than monthly, 2 = Monthly, 3 = Weekly, 4 = Daily or almost daily). All responses for alcohol frequency, quantity, and frequency of high episodic drinking were then added together and categorised based on the AUDIT-C scale, with scores of 0-4 indicating 'low risk', 5-7 indicating 'hazardous drinking', 8-10 indicating harmful drinking', and >10 indicating 'possible dependence' (National Institute for Health Clinical Excellence, 2022).

Physical Activity

Physical activity was assessed using the short version of the International Physical Activity Questionnaire (IPAQ), an internationally used self-report instrument for measuring physical activity. The IPAQ contains four questions that evaluate the volume and intensity of physical activity undertaken over the past week. These questions inquire about the duration (in days per week and minutes per day) and intensity (vigorous, moderate, and walking) of physical activity. Data on sedentary (sitting) behaviour was not collected. IPAQ is reliable and at least as valid as other physical activity measures for adults aged 18–65 (Craig et al., 2003).

Participants' responses to IPAQ were used to determine the total amount of physical activity completed in metabolic equivalents (MET) over a seven-day period. To do this, we first calculated the total number of minutes of activity across a 7-day period for each level of intensity (vigorous, moderate, and walking). Then, we multiplied the minutes of activity by the weighted MET estimate corresponding to the respective intensity level (see Supplementary Table 1 for the weights for each intensity category and the associated calculation; Hailey et al., 2023). Finally, we added the three values together to produce a measure of total physical activity per week (MET-min/wk).

Further, to maintain the reliability and validity of the questionnaire and allow for comparison with other studies that used IPAQ data, we categorised levels of physical activity in line with IPAQ guidelines (Hailey et al., 2023). Participants reporting implausible physical activity levels, that is those who reported a total activity time > 960 min (16 h) per day, assuming an average of 16 h of waking time, were assigned missing data flags. Those who reported < 10 min of activity per day were recoded to zero. Finally, data were truncated as in previous studies so that individuals exceeding 180 min in any intensity category were recoded as 180 min, permitting a maximum of 21 hours of activity in a week for each category (IPAQ Research Committee, 2005).

Thus, following the IPAQ scoring protocol, physical activity was categorised into three levels. 'Low activity' (<600 MET-min/week), moderate (600–3000 MET-min /week) or high (>3000 MET-min /week) (Pastuszak et al., 2014; Lear et al., 2017).

Sociodemographic variables

Data on sociodemographic variables was taken from the baseline Wave 7 dataset and included age, sex, ethnic group, educational attainment and monthly household income. *Age* was grouped in 10-year age bands, by starting with 16-24 year-olds and ending with the oldest category aged 65 and above.

Ethnic group was derived from the UKHLS measure for self-identified ethnic classification which allows participants to select the ethnic group to which they think they belong from a list of 18 categories for more details, see (Institute for Social and Economic Research and Essex, 2023). These categories are then collapsed into the following categories: White British; White Other (or Irish, Gypsy or Irish traveller; Any other White); Indian; Pakistani or Bangladeshi, Black (Black Caribbean/Black African), Other (all other ethnicities including mixed). In our grouping, the 5 minority groups (i.e. Pakistani, Bangladeshi, Indian, black African, and black Caribbean) are identified, not only because they are the largest minority groups in Great Britain, but also because the UKHLS over-samples these groups. To ensure consistency with previous studies using the UKHLS, we had initially grouped Pakistani, Bangladeshi and Indian together under the umbrella of 'South Asian', however we separated these since alcohol abstention rates (Pannu et al., 2009) and alcohol-related harms have been found to be distinctly different between Indians and South Asian ethnic minorities including Pakistanis and Bangladeshi in the UK (Hurcombe, Bayley and Goodman, 2010). Essentially, this collapsed version from the 18-group measure was used to ensure sufficient sample size in each ethnic group rather than to create homogenous categories of ethnic group.

Educational attainment was derived from the highest level of educational attainment variable and was recoded as: degree or higher, A-levels or equivalent; O-levels or equivalent; other educational qualifications; none.

For *monthly household income*, quintiles were calculated with weighted data: the UKHLS derived variable for net household income in the month before interview (which includes imputed values) was divided by the modified Organization for Economic Cooperation and Development (OECD) equivalence scale to account for the effects of household size and composition. All sociodemographic variables were converted into dummy variables.

Health outcomes measure

Disease data was used from the baseline Wave 7, and from the final Wave 13. Our key outcome variables were based on the respondents' self-reports of whether a medical doctor had ever told them they had any of a series of listed chronic conditions (see the first column in Table 1).

Researchers have measured multimorbidity using various methods (Wister et al., 2015). Given the lack of a uniform approach, we adopted the most widely cited and accepted definition of basic multimorbidity, which defines it as having two or more chronic conditions (Mercer et al., 2016). Consequently, we coded respondents as "yes" for multimorbidity if they had two or more conditions from the 16-condition list and as "no" otherwise.

We included health conditions that were assessed in both Wave 7 and Wave 13 to control for baseline diseases at Wave 7. It is important to note, however, that the methodology for measuring diseases and the format of questions posed to respondents about diseases underwent significant changes starting from Wave 10. Specifically, from Wave 10 onwards, continuing respondents were asked about their entire history of diagnosed health conditions, rather than only new conditions diagnosed since the previous wave, as was the case in earlier waves. Additionally, the measures for diseases became more granular; questions were expanded to include specifics about types of diabetes, cancer, and other conditions, allowing for a more

detailed assessment of health status. Consequently, for Wave 13, this meant that the prevalence of diseases could be directly ascertained from the responses at Wave 13. However, for continuing members from Wave 7, to assess whether a person had a disease, we needed to review their responses going backward from Wave 6 to Wave 1, where a diagnosis was last reported.

Consequently, the presence/absence of certain conditions (e.g., Cancer, Diabetes, and Arthritis) was incorporated in their aggregated form. This is because although information on these conditions was disaggregated in Wave 13, it was combined in Wave 7. Additionally, conditions not assessed in either wave (i.e. Clinical depression, HIV, hyperthyroidism) were excluded from the analysis. Moreover, since some of the listed conditions tend to co-occur or are known under different names by different cohorts of people, we also grouped conditions into six broader groups (see second column of Table 1). These groups were primarily derived from a recent paper by Fumagalli, Knies and Buck (2017). However, we made an adjustment to the grouping by Fumagalli, Knies and Buck (2017), whereby hypothyroidism, which had been paired with Type 1 diabetes and grouped under autoimmunity, was placed under 'other chronic conditions' since we did not have information on Type 1 diabetes at the baseline wave.

Finally, morbid obesity was included as a separate health outcome, which was measured as a binary variable in UKHLS as having a BMI of 40 or above.

Table 1. Chronic health conditions used to ascertain multimorbidity and their groupings

		Chronic health conditions	Grouping
1		Asthma	Respiratory
2		Emphysema	
3		Chronic bronchitis	
4		Congestive heart failure	Cardiovascular
5		Coronary heart disease	
6		Angina	
7		Heart attack or myocardial infarctions	
8		Stroke	
9	a)	Bowel/colorectal cancer*	Cancers
	b)	Lung cancer*	
	c)	Breast cancer (females only)*	
	d)	Prostate cancer (males only) *	
	e)	Liver cancer*	
	f)	Skin cancer or melanoma*	
	g)	Other cancer*	
10		High blood pressure/hypertension	Obesity related
11	a)	Diabetes†	
	b)	Gestational diabetes (females during pregnancy) †	
	c)	Other diabetes†	
12	a)	Osteoarthritis‡	Arthritis
	b)	Rheumatoid arthritis‡	
	c)	Other arthritis‡	
13		Hypothyroidism or an under-active thyroid	Other chronic
14		Any kind of liver condition	conditions
15		Multiple Sclerosis	
16		Other longstanding/chronic condition	

* Information on types of Cancer was only recorded in Wave 13, not in Wave 7
† Information on types of Diabetes types was only recorded in Wave 13, not in Wave 7
‡ Information on types of Arthritis was only recorded in Wave 13, not in Wave 7

Statistical Analysis

RMLCA was used to examine whether there were distinct classes of respondents who had similar patterns of SNAP behaviours over time. RMLCA adopts a probabilistic model-based approach that explicitly captures the number and composition of clusters, handles categorical data well, and allows for reliable interpretation and replication of patterns uncovered in the data. RMLCA was done using the MPlus v8.5 software and R version v4.0.3 (Muthén and Muthén, 2017a; Nylund-Gibson and Choi, 2018). A two-stage approach was used.

In the first stage, the optimal number of classes (i.e. clusters) was determined (Aim 1). For this, we entered data on the four health behaviours across four waves as independent data points to create a series of LCA models with increasing numbers of latent classes (i.e. clusters) until the model stopped converging. Model fit was then evaluated using several indices: Consistent Akaike's Information Criterion (CAIC), Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (aBIC), Approximate Weight of Evidence Criterion (AWE), Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT). Lower values of the BIC, CAIC, saBIC, and AWE indicate a better fitting model. However, when assessing the best model fit, particularly in large datasets with multiple indicators, additional classes can often lead to a decrease in the ICs — favouring the more complex model — until no further class can be added due to convergence issues (Nylund-Gibson and Choi, 2018). In such cases, the recommendation is to plot the ICs to seek a point of inflection or plateauing in the trend in the IC values as additional classes are added (Nylund-Gibson and Choi, 2018). For the likelihood-based tests such as the VLMR-LMR test, a p-value < 0.05 indicates that the model fit has not significantly improved compared to the model with one less class. Finally, classification diagnostic criteria such as entropy and the smallest average latent class posterior probability were also used to quantify how well the model separated individuals into distinct latent classes,

with values closer to 1 indicating distinct separation between classes for both measures. Missing data on the health behaviours were accounted for by Full Information Maximum Likelihood (Nylund-Gibson and Choi, 2018). Further, to assess the reliability of the class solution, we conducted a split-half replication where the sample was randomly split in half, and the aforementioned RMLCA was performed separately on these split samples to see if the solution for the full sample was replicated between these smaller splits (for details, see Supplementary Section A).

In the second stage, the associations between the latent classes identified in the first stage and socio-demographic characteristics and disease status were examined. This was done using the 3-step Bolck, Croon, and Hagenaars (BCH) method in which individuals are assigned to the class that they have the highest posterior probability of belonging to (Asparouhov and Muthén, 2014). To examine the association between socio-demographic characteristics and class membership (Aim 2), we regressed latent classes on socio-demographic variables in a series of multinomial logistic regressions, controlling for the potential inaccuracies in the class assignments, also known as classification errors, that are extracted as part of the 3-step method. To assess whether the prevalence of each health condition differed across classes (Aim 3), we regressed each health outcome on the latent classes using a binomial logistic regression that adjusted for classification errors reflect uncertainty in class assignment. For each health outcome, we conducted a series of binomial logistic regressions, that included: i) an unadjusted model, ii) a model adjusted for socio-demographic variables and, iii) a model adjusted for socio-demographic variables and presence of the health outcome at baseline (we did not control for morbid obesity at the baseline wave since obesity was only measured in Wave 13). We adopted this stepwise approach to examine how the adjustments influenced associations with

the disease outcomes and also because adjusting for baseline disease (which had a substantial amount of missing data) led to a decrease in sample size.

Then for each health outcome, in each set of these models, an omnibus Wald chi-square test was conducted to test for differences in disease prevalence across all latent classes (α=0.05). If significant, pairwise comparisons were conducted to test for differences in disease prevalence between each pair of classes. The significance level for pairwise comparisons was adjusted using the Bonferroni correction, set at a threshold level of $\alpha = 0.002$.

Additional post-hoc sub-group analysis

The complete sample analysis described above produced results that were difficult to interpret (see below). We therefore conducted further investigation by splitting our original dataset into two mutually exclusive samples. These samples were stratified by age, with one subset consisting of individuals aged 16-49 years (n=8110), and the other subset composed of individuals aged 50 years and older (n=9898). We chose to split the sample by age to assess whether the results of the complete sample analysis were attributable to patterns of multimorbidity in younger age groups that differed substantially from their older counterparts. This is plausible as chronic disease at early stages of the life course may have different causal processes to chronic disease at in later life (Wagner et al., 2024). For each split sample, we repeated the same analytical steps described above to address the same research objectives.

Results

Sample characteristics

The sample for our longitudinal analyses included 18,008 participants. The sample was predominantly White British (83.6%), with a slight female majority (56.9% females), and the average age was 50.8 years (range 16–95, s.d.=16.2). In total, 43.4% of the sample possessed a degree and 8.5% held no educational qualifications (see Table 2).

The sample's engagement in health risk behaviours across waves is shown in Table 3. The body system disorders with the highest prevalence were: obesity-related conditions (6.2%), Other conditions (4%), multimorbidity (3.4%), and arthritis (3.3%) (see Supplementary Table 2). These rates, although much lower than reported for the UK (Knies and Kumari, 2022), nonetheless reflect the prevalence rates observed in the complete UKHLS cross-sectional sample at Wave 13, which also reflected a relatively low prevalence for most conditions. Notably, the data from Wave 7 reveals a higher prevalence for most conditions, which likely mirrors the methodological changes in data collection and the approach to deriving these measures between the two waves.

Table 2. Baseline socio-demographic data of the final sample (n=18008)

	Socio-demographic characteristics at Wave 7	Study sample (n=18008) N (%)
Sex	Male	7760 (43.1)
	Female	10248 (56.9)
Ethnicity	White	15050 (83.6)
	White - Other	795 (4.4)
	Indian	574 (3.2)
	Pakistani and Bangladeshi	565 (3.1)
	Black (African/Caribbean)	414 (2.3)
	Other (including mixed ethnicities)	610 (3.4)
Education	Degree level or higher	7817 (43.4)
	A-levels or equivalent	3627 (20.1)
	O-levels or equivalent	3502 (19.4)
	Other educational qualifications	1525 (8.5)
	None	1537 (8.5)
Household income (in quintiles)	5th quintile (Highest)	3624 (20.1)
	4th quintile	3600 (20.0)
	3rd quintile	3601 (20.0)
	2nd quintile	3602 (20.0)
	1st quintile (Lowest)	3581 (19.9)
Age groups	16-24 years	1292 (7.2)
	25-34 years	1964 (10.9)
	35-44 years	2970 (16.5)
	45-54 years	3811 (21.2)
	55-64 years	3790 (21.0)
	65 years and older	4181 (23.2)

Table 3. Class-defining indicators (i.e. SNAP behaviours) among participants (n= 18008) included in the latent class analysis

		Wave 7	Wave 9	Wave 11	Wave 13
		N (%)	N (%)	N (%)	N (%)
Smoking	Non-smoker	15647 (86.9)	16039 (89.1)	16114 (89.5)	16321 (90.6)
	Smoker	2083 (11.6)	1889 (10.5)	1838 (10.2)	1647 (9.1)
	Missing	278 (1.5)	80 (0.4)	56 (0.3)	40 (0.2)
Fruit and vegetable intake	>= 5 portions/day	8011 (44.5)	9688 (53.8)	9622 (53.4)	9499 (52.7)
	< 5 portions/day	9997 (55.5)	8320 (46.2)	8386 (46.6)	8509 (47.3)
Alcohol consumption	Low-risk	9356 (52.0)	9440 (52.4)	9656 (53.6)	9512 (52.8)
	Hazardous	3923 (21.8)	3977 (22.1)	3550 (19.7)	3130 (17.4)
	Harmful	1039 (5.8)	1051 (5.8)	953 (5.3)	942 (5.2)
	Possible dependence	43 (0.2)	47 (0.3)	47 (0.3)	46 (0.3)
	Missing	3647 (20.3)	3493 (19.4)	3802 (21.1)	4378 (24.3)
Physical activity	High	4249 (23.6)	6016 (33.4)	5366 (29.8)	5069 (28.1)
	Medium	8299 (46.1)	7481 (41.5)	6829 (37.9)	7034 (39.1)
	Low	5175 (28.7)	4412 (24.5)	5726 (31.8)	5789 (32.1)
	Missing	285 (1.6)	99 (0.5)	87 (0.5)	116 (0.6)

Health behaviour clusters and within-cluster changes

The selection of the optimal latent class model was guided by examining key statistical criteria, classification diagnostics and interpretability considerations. After fitting models with one to nine latent classes, we compared models using traditional statistical methods including the VLMR-LMR test, log-likelihood, and fit information criteria and classification diagnostics like entropy to assess model fit (see Table 4). The VLMR-LMR test supported a five- and seven-class solution. Equally, the recommended range for entropy (value ≥0.8) and smallest class size (value ≥ 5%) was met by both the five-class and seven-class models (Nylund-Gibson and Choi, 2018). However, the values of the loglikelihood, BIC, SABIC, and CAIC continued to decrease as the number of classes increased, suggesting improved model fit (Chen et al., 2017). As the decline in information criteria plateaued around five classes (screeplot shown in Supplementary Figure 1), we decided to further compare the five-class and seven-class models using a sensitivity analysis on random split-halves of the data. This analysis also confirmed the robustness of the 5-class and 7-class solutions, as identical classes were uncovered in each split-half (full details can be found in Supplementary Section A).

Table 4. Model fit evaluation information for choosing a latent class model

K	LL	CAIC	BIC	SABIC	AWE	VLMR LRT p-value	Entropy	Smallest average latent class posterior probability	Smallest class size (%)
1	-203281.5	406619	406838	406749	406879.4	-	-	1	100
2	-189807.2	379728	380173	379992	380258.4	<0.001	0.974	0.984	14.6
3	-181641.8	363456	364126	363853	364255.3	0.002	0.876	0.916	13.9
4	-176776.9	353784	354681	354315	354853.1	0.003	0.821	0.879	13.6
5	-174867.4	350023	351146	350688	351361.8	**0.004**	0.805	0.798	13.6
6	-173497.1	347340	348689	348140	348948.9	0.105	0.790	0.763	10.8
7	**-172230.3**	**344865**	**346440**	**345798**	**346742.9**	**0.031**	**0.796**	**0.759**	**5.1**
8	-171261.4	342985	344786	344052	345132.8	0.234	0.777	0.756	5.1
9	-170382.3	341285	343312	342486	343702.2	0.080	0.777	0.751	4.6

Note: n=18008; K = number of classes (the nine-class model failed to converge); LL = model log likelihood; BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; p-value significance <0.05.

Given that these statistical indices did not uniformly point to a single model specification (and may over- or underestimate the number of classes present), the specific model estimates for both 5- and 7-class models were examined to select a final model specification based on the interpretability of the findings (i.e. whether it was possible to label the latent classes meaningfully), as well as theory and previous findings in the literature (e.g., Mawditt et al., 2016; Graham et al., 2016). Based on these criteria, the 7-class model was deemed the most appropriate for the data. In the 5-class model (see Supplementary Figure 3), distinct subgroups were identified, but all individuals categorised as smokers were grouped into one class, while hazardous drinkers formed another. While this aggregation simplifies the analysis, it sacrificed granularity, failing to capture potential heterogeneity in the sub-groups engaging in both drinking and smoking behaviours that has previously been observed in the UK (Mawditt et al., 2016; Meader et al., 2016) and even using the UKHLS (Graham et al., 2016; van den Broek, 2021). Each of these studies found a class exhibiting both smoking and drinking, and another class characterised by smokers engaging in multiple unhealthy behaviours. In contrast, the 7-class model disaggregated the smoking and drinking categories into multiple classes. This finer delineation allowed for the identification of subgroups characterised by varying levels and patterns of smoking and drinking behaviours. While model fit statistics may suggest multiple plausible solutions, prioritising interpretability enables researchers to extract meaningful insights and translate findings into practical implications for intervention and policy development (Sinha, Calfee and Delucchi, 2021). Thus, the 7-class solution was ultimately determined to be the best fitting model because it was the most conceptually interpretable and consistent with prior latent class studies of SNAP behaviours.

The seven distinct health-risk behaviour classes — hereafter referred to as clusters — were assigned labels by the research team. Figure 1 depicts these clusters and the average probability that individuals in each cluster engaged in the four health behaviours—smoking, alcohol

consumption, physical activity, and fruit and vegetable intake—over four distinct time points. Most behaviours were fairly stable over time, so unless highlighted below the behaviours that characterise each cluster were similar at each time point.

Figure 1. Seven-class model reflecting different clusters of health behaviour across time

Note. The x-axis lists each of the four behaviours – smoking, alcohol consumption, physical activity, and fruit and vegetable intake – across four time points. The y-axis provides the average probability for each of the indicators (i.e. four health behaviours) conditional on membership in a given class (i.e. cluster).

Participants in the *Overall low-risk* cluster (20% of the sample) maintained adequate (≥5 per day) fruit and vegetable intake and high levels of physical activity across time. They also had a consistently high probability of drinking at low-risk levels, and the proportion of people consuming alcohol at hazardous levels fell across waves.

Participants in the *Insufficiently active* cluster (18.3%) displayed low-to-medium levels of physical activity. However, they consumed the recommended portions of fruit and vegetables, and had low-risk alcohol intake.

The largest proportion of participants fell in the *Poor diet, Insufficiently active* cluster (23.2%). Participants in this cluster exhibited patterns of smoking and alcohol consumption similar to those in the *Overall low-risk* cluster. However, participants in the *Poor diet, Insufficiently active* cluster had a high probability of not maintaining adequate fruit and vegetable intake and displayed low-medium levels of physical activity.

Participants in the *Hazardous and harmful drinkers* cluster (10.7%) had a high likelihood of consuming alcohol at hazardous or harmful levels. They consumed the recommended portions of fruit and vegetables and had the same probability of engaging in physical activity across all three levels (i.e. high, medium, low).

Participants in the *Hazardous drinkers, Insufficiently active and Poor diet* cluster (14%) had a similar profile to the *Poor diet, Insufficiently active* cluster, except they had a consistently high probability of consuming alcohol at hazardous levels.

Participants characterised as *Smokers and drinkers* (5.1%) had a high probability of smoking, increasing over time then declining over the last wave, adequate fruit and vegetable intake and low levels of physical activity. Notably, participants in this cluster exhibited patterns of alcohol consumption and physical activity similar to those in the *Hazardous and harmful drinkers* cluster.

Finally, participants in the *Smokers* cluster (8.7%) had a high probability of smoking, albeit declining over time, medium to high levels of physical activity, and adequate fruit and vegetable intake.

Clusters and socio-demographic characteristics

Table 5 displays the socio-demographic composition of participants in each of the identified clusters and the results from the adjusted multinomial logistic regressions that examined associations between each socio-demographic characteristic and cluster membership.

Compared to participants in the *Overall low-risk* cluster (which served as the reference group), participants characterised as *Smokers and Drinkers* and *Smokers* were about two times more likely to be aged 25-44 years than to fall in the youngest age group. Conversely, participants in the *Overall low-risk* cluster were more likely to be 65 years and older than in the youngest age group compared to all other clusters except *Poor diet, Insufficiently active*. In turn, participants characterised as *Poor diet, Insufficiently active,* were around 1.7 times more likely to fall in age group 55-64 years, compared to the *Overall low-risk* cluster.

Women made up around two-thirds of the *Overall low-risk* and *Unhealthy diet, Insufficiently active* clusters, while men comprised two-thirds of the *Hazardous and harmful drinkers*. Notably, the *Overall low-risk* cluster (reference group) had a higher likelihood of female membership compared to all other clusters.

In keeping with the ethnic distribution of UKHLS, all clusters were predominantly White British (which served as the reference category). Yet, participants characterised as *Poor diet, Insufficiently active* were 1.5 times more likely to belong to other White ethnicities compared to the *Overall low-risk* cluster. In contrast, compared to the *Overall low-risk* cluster, other clusters were less likely belong to ethnic minorities such as Indian, Pakistani or Bangladeshi, or Black as well as ethnicities categorised as other (which included mixed ethnicities).

Participants in two clusters – *Poor diet, Insufficiently active* and *Hazardous drinkers, Insufficiently active with Poor diet* – had the highest proportion (approximately 50%) of individuals educated to degree level or higher. Compared to the *Overall low-risk* cluster, these two clusters had a lower likelihood of their highest qualification being at any other levels, such as A-level, O-level, other educational qualifications, or no qualifications. Conversely, participants characterised as *Smokers and Drinkers* and *Smokers* were nearly twice as likely to have their highest educational qualification being at other levels (i.e. qualifications other than at degree levels), compared *Overall low-risk* cluster. Notably, *Smokers* were 2.5 more likely to have no educational qualifications compared to the *Overall low-risk* cluster.

Compared to the *Overall low-risk* cluster, three clusters—namely, '*Poor diet, Insufficiently active; Hazardous and harmful drinkers; and Hazardous drinkers, Insufficiently active with Poor diet*—had a higher likelihood of belonging to the highest quintile of household income, as opposed to any other quintiles. Conversely, participants characterised as *Smokers* were two times more likely to belong to the fourth (i.e. second highest) quintile of household income compared to the *Overall low-risk* cluster.

Table 5. Demographics and Odds ratios from Multinomial Logistic regressions examining the association between socio-demographic predictors and cluster membership.

	Overall low-risk (Ref)		Insufficiently Active	Poor diet, Insufficiently active		Hazardous and harmful drinkers		Hazardous drinkers, Insufficiently active with Unhealthy diet		Smokers and Drinkers		Smokers
	19.97%		18.29%	23.22%		10.72%		14.03%		5.09%		8.68%
			OR [95% C.I]	OR [95% C.I]		OR [95% C.I]		OR [95% C.I]		OR [95% C.I]		OR [95% C.I]
Age group												
16-24 (Ref)	10.2%	15.7%	Ref.	8.0%	Ref.	14.6%	Ref.	14.9%	Ref.	16.3%	Ref.	11.4%	Ref.
25-34	10.2%	16.9%	0.96 [0.62, 1.49]	10.0%	0.95 [0.61, 1.47]	11.7%	0.76 [0.45, 1.31]	12.8%	0.65 [0.42, 1.01]	20.2%	1.36 [0.75, 2.46]	16.8%	**1.97 [1.16, 3.34]**
35-44	10.7%	12.4%	0.70 [0.47, 1.06]	15.0%	1.45 [0.98, 2.15]	19.1%	1.23 [0.78, 1.95]	18.3%	0.96 [0.64, 1.44]	16.2%	1.07 [0.62, 1.85]	17.3%	**1.84 [1.12, 3.02]**
45-54	14.8%	18.5%	0.70 [0.47, 1.04]	19.4%	1.31 [0.90, 1.92]	20.6%	0.83 [0.54, 1.30]	23.6%	0.80 [0.54, 1.17]	22.5%	0.89 [0.53, 1.49]	20.2%	1.34 [0.84, 2.16]
55-64	14.0%	15.6%	**0.64 [0.43, 0.95]**	21.0%	**1.65 [1.13, 2.42]**	21.9%	0.91 [0.59, 1.42]	19.0%	**0.67 [0.45, 0.99]**	17.0%	0.69 [0.41, 1.17]	19.5%	1.27 [0.79, 2.04]
65 and above	40.1%	20.9%	**0.35 [0.24, 0.52]**	26.6%	0.97 [0.67, 1.40]	12.1%	**0.21 [0.13, 0.33]**	11.4%	**0.18 [0.12, 0.27]**	7.8%	**0.12 [0.07, 0.21]**	14.8%	**0.27 [0.17, 0.45]**
Sex													
Male (Ref)	31.7%	54.1%	Ref.	35.8%	Ref.	66.0%	Ref.	58.8%	Ref.	59.2%	Ref.	38.3%	Ref.
Female	68.3%	45.9%	**0.36 [0.30, 0.44]**	64.2%	**0.83 [0.70, 0.98]**	34.0%	**0.21 [0.17, 0.27]**	41.2%	**0.30 [0.25, 0.36]**	40.8%	**0.27 [0.21, 0.36]**	61.7%	**0.66 [0.53, 0.83]**
Ethnic group													
White - British (Ref)	86.0%	89.6%	Ref.	88.4%	Ref.	92.6%	Ref.	94.9%	Ref.	91.6%	Ref.	87.9%	Ref.
White - Other	2.5%	4.1%	1.34 [0.80, 2.26]	4.7%	**1.56 [1.04, 2.34]**	3.2%	0.96 [0.51, 1.82]	2.7%	0.80 [0.47, 1.36]	2.4%	0.75 [0.37, 1.53]	4.3%	1.39 [0.83, 2.35]
Indian	3.2%	1.7%	**0.30 [0.17, 0.53]**	1.9%	**0.43 [0.28, 0.65]**	0.9%	**0.14 [0.06, 0.32]**	0.6%	**0.09 [0.04, 0.24]**	0.9%	**0.15 [0.06, 0.38]**	0.8%	**0.18 [0.09, 0.37]**
Pakistani or Bangladeshi	2.8%	1.3%	**0.29 [0.16, 0.52]**	0.6%	**0.22 [0.12, 0.41]**	1.3%	**0.28 [0.13, 0.60]**	0.1%	**0.03 [0.01, 0.19]**	0.9%	**0.15 [0.06, 0.40]**	1.5%	**0.31 [0.18, 0.53]**
Black - African/Caribbean	2.4%	1.7%	0.53 [0.28, 1.00]	1.3%	**0.45 [0.27, 0.74]**	0.4%	**0.10 [0.02, 0.44]**	0.5%	**0.16 [0.05, 0.52]**	0.4%	**0.09 [0.02, 0.36]**	1.9%	**0.55 [0.32, 0.97]**
Other	3.1%	1.6%	**0.36 [0.20, 0.67]**	3.1%	0.76 [0.49, 1.17]	1.6%	**0.31 [0.14, 0.67]**	1.2%	**0.21 [0.10, 0.43]**	3.8%	0.81 [0.36, 1.79]	3.6%	0.93 [0.53, 1.62]
Educational Qualification													
Degree level or higher (Ref)	29.1%	39.3%	Ref.	50.9%	Ref.	37.4%	Ref.	50.2%	Ref.	23.0%	Ref.	18.7%	Ref.
A-levels or equivalent	17.2%	23.6%	1.03 [0.80, 1.33]	19.0%	**0.76 [0.61, 0.94]**	25.0%	1.24 [0.94, 1.63]	24.2%	0.94 [0.74, 1.20]	26.8%	**1.90 [1.29, 2.79]**	21.2%	**1.84 [1.34, 2.54]**
O-levels or equivalent	18.4%	22.7%	1.07 [0.84, 1.37]	16.0%	**0.58 [0.47, 0.72]**	24.3%	**1.33 [1.01, 1.74]**	16.7%	**0.76 [0.59, 0.97]**	31.9%	**2.52 [1.76, 3.61]**	27.9%	**2.38 [1.78, 3.19]**
Others	13.5%	8.0%	**0.60 [0.42, 0.84]**	8.6%	**0.44 [0.34, 0.57]**	7.3%	0.72 [0.49, 1.04]	5.6%	**0.46 [0.33, 0.66]**	13.3%	**2.04 [1.28, 3.25]**	12.5%	**1.90 [1.30, 2.76]**
None	21.8%	6.4%	**0.39 [0.27, 0.55]**	5.5%	**0.20 [0.15, 0.26]**	6.0%	**0.50 [0.33, 0.75]**	3.3%	**0.26 [0.17, 0.40]**	5.0%	0.66 [0.34, 1.28]	19.7%	**2.46 [1.76, 3.45]**
Household income													

Highest Quintile (Ref)	9.8%	17.2%	Ref.	21.3%	Ref.	22.5%	Ref.	29.9%	Ref.	12.8%	Ref.	5.4%	Ref.
4th Quintile	14.6%	19.6%	**0.58 [0.43, 0.80]**	22.2%	**0.43 [0.33, 0.56]**	19.4%	**0.37 [0.26, 0.52]**	25.1%	**0.20 [0.15, 0.28]**	13.0%	0.92 [0.58, 1.43]	11.2%	**2.21 [1.49, 3.28]**
3rd Quintile	20.3%	20.5%	**0.50 [0.37, 0.68]**	20.3%	**0.50 [0.39, 0.65]**	20.4%	**0.34 [0.24, 0.48]**	19.9%	**0.23 [0.17, 0.31]**	16.9%	0.72 [0.47, 1.12]	19.2%	1.47 [0.99, 2.18]
2nd Quintile	27.6%	20.8%	**0.58 [0.43, 0.78]**	20.6%	**0.55 [0.43, 0.71]**	19.6%	**0.42 [0.30, 0.59]**	14.4%	**0.34 [0.26, 0.46]**	27.4%	**0.56 [0.35, 0.89]**	26.1%	1.44 [0.96, 2.16]
Lowest Quintile	27.7%	21.9%	0.74 [0.54, 1.02]	15.6%	**0.75 [0.58, 0.97]**	18.1%	**0.55 [0.39, 0.76]**	10.7%	**0.57 [0.43, 0.75]**	29.9%	**0.59 [0.37, 0.96]**	38.1%	1.21 [0.78, 1.87]

Note**: Bold values** show significant odds ratios at significance level p=0.05.

Health behaviour clusters and disease status

After controlling for both sociodemographic characteristics and respective diseases at baseline, only the prevalence of multimorbidity, respiratory disorders, arthritis, cardiovascular diseases and morbid obesity differed significantly across clusters — omnibus Wald test χ^2(df = 6) multimorbidity = 19.47, p = 0.003; χ^2(df = 6) respiratory disorders = 21.09, p=0.002; χ^2(df = 6) Arthritis = 17.21, p=0.009; and χ^2(df = 6) cardiovascular diseases = 12.62, p =0.050; χ^2(df = 6) morbid obesity =84.12, p<0.001, respectively. To further investigate which clusters differed in the proportions of participants with each disease profile, we conducted pairwise comparisons between clusters, with results adjusted for disease status and socio-demographic variables, for multimorbidity, respiratory disorders, arthritis, and cardiovascular diseases, as shown in Table 6.

An initial assessment of the data revealed an unexpected trend: Participants characterised as *Overall low-risk* had a higher prevalence of multimorbidity compared to participants in the *Hazardous drinkers, Insufficiently active with Poor diet* clusters. Participants characterised as *Overall low-risk* also had a higher prevalence of morbid obesity compared to all other clusters except the cluster labelled *Smokers and Drinkers*. Notably, the higher prevalence of multimorbidity in the *Overall low-risk* persisted after adjusting for sociodemographic factors. This unexpected finding warrants a more thorough examination and will be further explored in the discussion section.

Similarly, participants characterised as *Smokers* had a higher prevalence of respiratory disorders than the *Insufficiently Active* cluster. *Smoker and Drinkers* had a similarly high prevalence of respiratory disorders to *Smokers*.

Although we found significant differences across clusters in the prevalence of arthritis and cardiovascular diseases, the pairwise differences were not significant at the Bonferroni corrected level. However, we did observe that participants characterised as *Smokers* and *Overall low-risk* had a higher prevalence of both arthritis and cardiovascular diseases compared to other clusters.

Table 6. Pairwise comparisons of disease status prevalence across clusters

Disease outcomes (Wave 13)	Overall low-risk	Insufficiently Active	Poor Diet, Insufficiently active	Hazardous and harmful drinkers	Hazardous drinkers, Insufficiently active with Poor diet	Smokers and Drinkers	Smokers	Wald's omnibus test χ^2(df = 6)	Wald's omnibus p-value	Sample size
	19.97%	**18.29%**	**23.22%**	**10.72%**	**14.03%**	**5.09%**	**8.68%**			
Multimorbidity (>= 2 chronic conditions)										
Unadjusted	0.066	0.017	0.027	0.026	0.01	0.024	0.049	79.02	**p<0.001**	n=18008
Adjusted for socio-demographic covariates	0.061	0.02	0.027	0.027	0.012	0.025	0.048	33.47	**p<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.063 [a]	0.022	0.03	0.027	0.012 [a]	0.026	0.052	19.47	**p=0.003**	n=14489
Cancers (All types)										
Unadjusted	0.016	0.009	0.014	0.007	0.007	0.01	0.015	8.94	p= 0.177	n=18008
Adjusted for socio-demographic covariates	0.015	0.01	0.013	0.008	0.008	0.01	0.014	4.70	p=0.583	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.015	0.011	0.014	0.007	0.008	0.01	0.018	0.06	p=1.000	n=13368
Obesity-related (Diabetes, High Blood pressure)										
Unadjusted	0.092	0.038	0.058	0.059	0.035	0.048	0.063	48.15	**p<0.001**	n=18008
Adjusted for socio-demographic covariates	0.087	0.04	0.057	0.059	0.037	0.049	0.062	20.20	**p=0.003**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.089	0.047	0.06	0.063	0.037	0.047	0.064	8.70	p=0.1911	n=13815
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)										
Unadjusted	0.037	0.009	0.022	0.017	0.017	0.048	0.059	54.27	**p<0.001**	n=18008
Adjusted for socio-demographic covariates	0.035	0.01	0.022	0.017	0.017	0.049	0.057	37.20	**p<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.04	0.012 [a]	0.023	0.013	0.02	0.055	0.055 [a]	21.09	**p=0.002**	n=13418
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)										
Unadjusted	0.054	0.016	0.03	0.033	0.013	0.019	0.049	53.34	**p<0.001**	n=18008
Adjusted for socio-demographic covariates	0.05	0.018	0.029	0.035	0.014	0.021	0.047	26.94	**p<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.052	0.02	0.031	0.039	0.015	0.023	0.052	17.21	**p=0.009**	n=13619
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)										
Unadjusted	0.038	0.015	0.01	0.021	0.005	0.012	0.032	47.85	**p<0.001**	n=18008
Adjusted for socio-demographic covariates	0.035	0.016	0.011	0.021	0.005	0.013	0.031	31.22	**p<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.033	0.019	0.011	0.02	0.008	0.017	0.03	12.62	**p=0.050**	n=13390
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)										
Unadjusted	0.066	0.036	0.039	0.021	0.026	0.021	0.055	44.80	**p<0.001**	n=18008
Adjusted for socio-demographic covariates	0.064	0.038	0.038	0.023	0.026	0.021	0.055	23.10	**p=0.001**	n=18008

*Adjusted for socio-demographic covariates and baseline disease	0.063	0.041	0.037	0.026	0.028	0.02	0.06	14.29	p=0.027	n=13390
Morbid Obesity (BMI ≥40)										
Unadjusted	0.076	0.018	0.031	0.029	0.018	0.036	0.051	78.26	**p<0.001**	n=18008
Adjusted for socio-demographic covariates	0.077 **a,b,c,d,e**	0.018 **a**	0.031 **b**	0.029 **d**	0.018 **c**	0.036	0.05 **e**	84.12	**p<0.001**	n=18008

Note: Socio-demographic covariates adjusted for include: sex, age, own occupation, education level and monthly household income quintile.
a,b,c,d,e indicates statistically significant pairwise differences in disease prevalence at Bonferroni corrected significance level of p=0.002.
Bold values indicate a statistically significant omnibus wald's test at significance level of p=0.05.

Results of the age-stratified subsamples

The finding that the *Overall low-risk cluster* exhibited higher multimorbidity prevalence presented an interpretive challenge. To assess this further, we split our original dataset into two mutually exclusive age-stratified samples. One subset consisted of individuals aged 16-49 years (n=8110), and the other subset composed of individuals aged 50 years and older (n=9898). This age division aimed to discern if the observed multimorbidity patterns in the aggregate data were a reflection of age-specific disease trends, in line with the hypothesis that disease aetiology in younger populations might diverge from that in the older adults (Wagner et al., 2024). Subsequently, we repeated the same analytical steps described above to address the same research objectives for each split-sample.

The supplementary material (sections B and C) provides a comprehensive breakdown of the results obtained from age-stratified subsamples. Specifically, within the older subsample (aged 50 years and over), a five-class model emerged as the best fit for the data, whereas a four-class model demonstrated optimal fit for the younger subsample (aged 49 years and younger). As anticipated, the overall prevalence of all examined conditions was lower in the younger subsample compared to the older one.

Both subsamples included four distinct clusters with similar behavioural profiles: i) Overall low-risk class, ii) Poor diet, Insufficiently Active, iii) Hazardous drinkers, and iv) Smokers. Notably, the older subsample revealed an additional cluster that we labelled Insufficiently Active.

The younger subsample displayed minimal differences between clusters in most of the health conditions due to their low overall prevalence. Significant between-cluster differences were only found for morbid obesity and respiratory disorders, in line with the differences in the complete sample. Similarly, the older subsample exhibited a higher prevalence of most health conditions in the clusters labelled *Overall low-risk* and *Smokers*, mirroring trends observed in the complete sample.

In sum, the age-stratified analysis broadly aligned with the findings in the complete sample (i.e. higher prevalence of most health conditions studied in the clusters labelled *Overall low-risk* and *Smokers*).

Discussion

This study investigated the relationship between clusters of health-risk behaviours over time and multimorbidity in a representative sample of adults (aged 16 and over) in the UK. We identified seven clusters representing distinct profiles of health behaviour. Although heterogeneity in the literature, with regards to lifestyle factors assessed and measures used, makes it difficult to draw firm conclusions, we noted striking similarities in the risk behaviour clusters identified in our study and those found in previous studies on adult populations with similarly large age ranges as ours in the UK (Tegegne, Islam and Maddison, 2022; Birch et al., 2019), Ireland (Conry et al., 2011) and Western Australia (French, Rosenberg and Knuiman, 2008). This included the presence of an overall low-risk cluster, a physically inactive cluster with poor diet, a cluster characterised primarily by heavy alcohol consumption, and a cluster with smokers exhibiting additional risk behaviour(s). Notably, these four core clusters were also represented in the study on older adults using the ELSA study (Suhag, Webb and Holmes, 2024). Nevertheless, the ELSA study had only one cluster with current smokers, while the present study comprised two. This observation aligns with recent trends suggesting a steady decline in smoking with increases in age beyond 25 years, thus explaining the smaller proportion of current smokers in the overall sample of older adults examined (Opazo Breton et al., 2022).

Notably, the present study advances existing research by using longitudinal data to not only examine whether distinct clusters of health behaviours are found in a nationally-representative sample of adults but also whether patterns of behaviour within each cluster change over time. Consistent with a previous study on older adults in England (Suhag, Webb and Holmes, 2024), we found that the patterns of behaviour within the clusters were largely stable over time. This observation is consistent with prior evidence suggesting that risk behaviours are relatively

stable during mid-adulthood (Mawditt et al., 2016; Benzies, Wångby and Bergman, 2008) and older age (Botoseneanu and Liang, 2012), in contrast to adolescence and young adulthood (Daw, Margolis and Wright, 2017; Fromme, Corbin and Kruse, 2008). However, there were some exceptions to this trend of stability: The proportion of people consuming alcohol at hazardous levels fell across waves in the *Overall low-risk* and *Poor diet, Insufficiently active* clusters. This finding is noteworthy considering that these clusters had the highest proportion of older individuals (i.e. approximately half of the individuals in these two clusters were aged 55 years and older) and exhibited a higher proportion of females compared to other clusters, with a two-thirds female majority. This pattern of diminishing alcohol consumption in older demographics, particularly among women, corresponds with findings from earlier studies on alcohol consumption patterns across different ages in the UK (Britton et al., 2015). In contrast, the proportion of current smokers gradually rose over time to then decline in the last wave in the *Smokers and Drinkers* cluster, while the proportion of smokers consistently fell over time the *Smokers* cluster. The declining rates of current smokers over time align with a wealth of evidence pointing to a steady reduction in smoking prevalence in the UK among all age groups older than 25 (McGeoch et al., 2023). Importantly, the trends over time exhibited in these exception cases were also comparable across the current study and the study by Suhag, Webb and Holmes (2024). This similarity suggests a level of consistency and reliability in the observed behavioural dynamics across the two samples, despite their varying age demographics, likely mirroring broader cultural trends within the UK (Opazo Breton et al., 2022; Wadd and Papadopoulos, 2014; Whitaker et al., 2020).

The identified clusters also differed in their sociodemographic profiles. The *Smokers and Drinkers* cluster had the highest proportion of younger individuals of all clusters, with over half of the population aged 44 years and younger, and a two-thirds male majority, consistent

with previous research indicating a clustering of alcohol consumption and smoking among males of working age (Noble et al., 2015). Similarly, the three clusters of hazardous and harmful drinkers were predominantly male, in line with alcohol consumption patterns observed in the UK (Meier et al., 2021). Two clusters namely, *Overall low-risk* and *Poor diet, Insufficiently active,* were characterised by older individuals — approximately half of the population was aged 55 years or older — and primarily comprised women. Notably, the socio-demographic and behavioural profile of the *Poor diet, Insufficiently active* cluster, which encompassed older women from advantaged socio-economic groups characterised by higher levels of education and wealth, appear to closely resemble those of the 'Healthy' or 'Safe' clusters identified in prior studies (Conry et al., 2011; French, Rosenberg and Knuiman, 2008; Suhag, Webb and Holmes, 2024). We also noted differences in the ethnic distribution of the clusters, whereby the minority groups examined in the UKHLS (i.e. Indian, Pakistani, Bangladeshi, Black African and Black Caribbean) were more likely to fall in the *Overall low-risk cluster*. Other UK studies have pointed to healthier behaviours among minority ethnic groups (Graham et al., 2016; Lawder et al., 2010). Equally, two clusters, namely the *Overall low-risk* and *Smokers* were least likely to belong to the highest quintile of wealth and most likely to hold no educational qualifications. While the absence of risk behaviour clustering and socioeconomic disadvantage observed in the *Overall low-risk* is harder to explain, it is worth noting that the socioeconomic disadvantage and female majority within the cluster labelled *Smokers* mirrors findings from a recent study on older adults in England (Suhag, Webb and Holmes, 2024). The sociodemographic profile of *Smokers* may be explained by the higher prevalence of smoking among disadvantaged groups, coupled with the increased exposure to tobacco's harms among disadvantaged smokers (Hiscock et al., 2012). As a consequence, the higher proportion of females in this cluster may indicate survivorship bias, as smoking is the

leading cause of lung cancer deaths, but lung cancer occurs less frequently and has a better prognosis in women (Barta, Powell and Wisnivesky, 2019).

Importantly, identified clusters also differed in their disease status, with some findings more predictable than others. For instance, *Smokers* had a similar prevalence of respiratory disorders to that of *Smokers and Drinkers,* and had a higher prevalence than the *Insufficiently Active* cluster, aligning with established research on the topic (Jayes et al., 2016). Conversely, participants characterised as *Overall low-risk* had a higher prevalence of morbid obesity compared to all other clusters except the cluster labelled *Smokers and Drinkers.* The *Overall low-risk* also had the highest prevalence of multimorbidity among all clusters. Since morbid obesity is itself a key risk factor for multimorbidity, the co-occurrence of these outcomes in the *Overall low-risk* is unsurprising (Shan, Yin and Panuthai, 2024). Yet, interpreting these results solely through the lens of risk behaviour profiles may present a limited picture, particularly regarding dietary intake. The assessment of diet in the UKHLS primarily relies on weekly fruit and vegetable intake, which, while suitable for screening purposes, may not fully capture disease-attributable risks of diet (Ocean, Howley and Ensor, 2019). This limitation arises because diet quality encompasses both nutrient density and energy intake, and a sufficient intake of nutrient-dense foods (e.g., fruits and vegetables), may contribute little to energy intake (Wirt and Collins, 2009). Moreover, the omission of energy-dense, nutrient-poor ultra-processed foods that make up more than half of the total dietary energy consumed in the UK, leaves gaps in our comprehension of diet-related risks (Rauber et al., 2018).

Nevertheless, as the *Overall low-risk* cluster comprised primarily women, were older, and exhibited notable socioeconomic disadvantages (i.e. more likely to fall in lower quintiles of wealth and lower education levels), our findings are consistent with a recent study on lifestyle

risk factors for multimorbidity which found that among women, none of the unhealthy SNAP behaviour had an independent and statistically significant association with multimorbidity, but being older, less educated, and overweight was associated with multimorbidity (de Almeida et al., 2020). However, the differences in multimorbidity prevalence remained after controlling for socio-demographic characteristics. This presented an interpretative challenge and is discussed in further detail in the limitations section. We also found the prevalence of arthritis and cardiovascular diseases to be higher among *Smokers* and the *Overall low-risk cluster*. While smoking behaviour explains the elevated cardiovascular disease prevalence among *Smokers* (Burns, 2003), the same cannot be said for the *Overall low-risk.* Yet, when considering the relatively high prevalence of morbid obesity in both these clusters in combination with socioeconomic disadvantage, both of which greatly increases the risk of eventually developing arthritis (Pandey et al., 2013) as well as cardiovascular disease (Van Gaal, Mertens and De Block, 2006), the high prevalence of these conditions in the clusters can be partially explained. Finally, the *Hazardous drinkers, Insufficiently active with Poor diet* cluster had a lower prevalence of all health conditions studied. Of note, this cluster had a distinctive sociodemographic profile: most likely to belong to highest quintile of wealth, be educated to the highest level and least likely to belong to a minority ethnic group. Overall, these findings align with extensive evidence on the influence of socioeconomic status on the development and prevalence of chronic illness and multimorbidity, suggesting that individuals from more socioeconomically deprived backgrounds are inherently more vulnerable to the adverse effects of unhealthy lifestyles than their more affluent counterparts (Mair and Jani, 2020; Foster et al., 2018).

Strengths and limitations

A number of strengths distinguish this study. This study is the first to explore the relationship between longitudinal clusters of multiple health-risk behaviours and multimorbidity using a nationally-representative sample of adults across a broad age range, starting from 16 years and older. It employs data from the UKHLS, which involves individuals from across the UK and therefore allows a high level of generalisability. It uses a robust, model-based, probabilistic approach (namely, RMLCA), that has several advantages over other traditional methods used to study lifestyle risk behaviours (Schuit et al., 2002; Dumith et al., 2012) in that it: i) allows for the examination of multiple behaviours simultaneously, ii) demonstrates stable results in split-half replication, and iii) can be reproduced and validated since diagnostic criteria are accessible and programming codes can be made available upon request.

While our study benefits from several strengths, our findings should be considered with the following limitations in mind. Firstly, relying on self-reported measures of chronic diseases and risk behaviours may introduce biases, potentially leading to underestimation (e.g., Meier et al., 2013) or overestimation (e.g., Cleland et al., 2018) of their prevalence. Secondly, certain risk behaviours are inadequately specified; for instance, the dietary measure utilised in the UKHLS focuses solely on fruit and vegetable intake, neglecting other dietary components. Thirdly, our study primarily explores the prevalence of health conditions rather than their incidence rates. Finally, the absence of a temporal gap between exposure and outcome means we cannot draw causal deductions.

Although we controlled for various individual-level socio-demographic covariates associated with multimorbidity, we were unable to account for environmental and area-level factors that might influence health outcomes (Ingram et al., 2021). Similarly, evidence suggests that

cultural norms, social networks, and community support systems, which were not included in our model, can significantly influence health outcomes and disparities in these outcomes, even after demographic variables are accounted for (Chuang, Chuang and Yang, 2013). These unaccounted factors may explain the persistent differences in disease prevalence between clusters, despite controlling for socio-demographic covariates.

Practical implications and future research

By examining how behavioural risk factors cluster in the context of chronic disease in a nationally representative sample, the present study sheds light on the relationship between health behaviours, socioeconomic disparities, and multimorbidity among adults in the United Kingdom. Further, the conceptual replication of Suhag, Webb, and Holmes' (2024) study within the same cultural milieu—while extending it to a wider age spectrum and observing similarities in cluster patterns—bolsters the reliability of the study's results. These findings hold significant implications for the design and implementation of health promotion initiatives.

Two key high-risk subgroups, characterised by poorer health outcomes, emerged from our analysis: the *Overall low-risk cluster* and *Smokers*. These groups not only exhibited poorer health outcomes but were also socioeconomically disadvantaged. Interestingly, while we did not observe a straightforward linear association between the number of risky behaviours and disease prevalence, we did uncover that clusters engaging in multiple risky behaviours but with favourable socioeconomic profiles tended to have better health outcomes, mirroring the findings in the study by Suhag, Webb and Holmes (2024).

Similarly, our findings underscore the enduring influence of socioeconomic background on health behaviours and disease outcomes, whereby the ostensibly healthy behaviours of the *Overall low-risk* did not seem to curtail their high prevalence of multiple chronic diseases. Moreover, we highlight the critical role of ethnic background. Consistent with previous research (Graham et al., 2016; Lawder et al., 2010), our study found that the cluster with the healthiest risk behaviour profile, was more likely to comprise ethnic minorities, but had worse health outcomes. Thus, tailoring chronic disease prevention efforts to specific population subgroups and their corresponding risk factor profiles, as elucidated in our analysis, holds significant promise.

However, our study also cautions against a one-size-fits-all approach. The discovery that the *Overall low-risk cluster*, characterised by the absence of risky behaviours, experienced worse outcomes underscores the need to consider broader social determinants of health. It suggests that a narrow focus on behavioural clustering, without accounting for the complex interplay of social forces shaping individual behaviours, may inadvertently lead to counterproductive outcomes. Our findings advocate for a nuanced understanding of risks behaviours within a social context. While measurement and monitoring of behavioural clustering can indeed inform targeted health promotion strategies, it must be complemented by a recognition of the structural influences on individual choices. By integrating these insights with big data from healthcare providers and individual risk assessments, we can pave the way for more effective and comprehensive healthcare interventions tailored to the diverse needs of high-risk population subgroups.

Conclusions

In conclusion, the present research identified seven distinct clusters among adults, each characterised by a unique pattern of risk behaviours, socio-demographic characteristics, and multimorbidity prevalence within a nationally-representative sample. Notably, our investigation is among the first to explore the relationship between multiple risk behaviours (examined simultaneously over time) and multimorbidity in adults, filling a notable gap in the existing literature. However, the paucity of comparable nationally-representative longitudinal studies limits direct comparison of our findings.

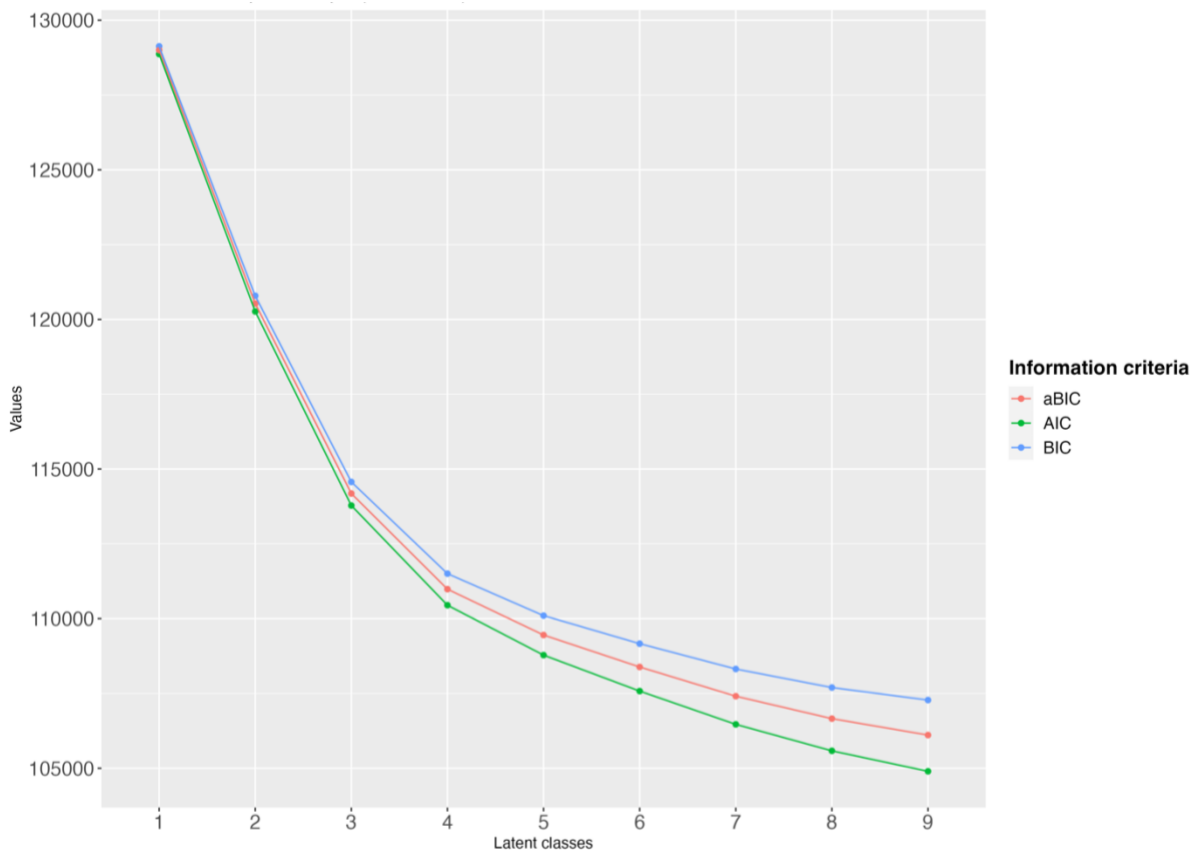
Our research revealed that simply examining the number or combination of risk behaviours falls short in explaining the variations in disease outcomes across different clusters. Instead, we observed a consistent presence of socioeconomic disadvantage and female gender among clusters with poorer health outcomes. This underscores the crucial role of social determinants in shaping health disparities and underscores the necessity of considering these factors in policy formulation. Furthermore, our findings underscore a notable trend: health-risk behaviours tend to improve with age, suggesting potential avenues for targeted interventions across the lifespan. Future investigations should examine the associations of risk behaviour clusters and multimorbidity, using a more comprehensive measure of diet and measuring incidence in addition to prevalence, where possible. Finally, research on health behaviour patterns should focus on linking identified clusters to other important clinical outcomes (e.g., inflammatory biomarkers) in order to identify vulnerable groups and to allow for individualised patient-centred primary prevention programs.

Supplementary Material

Supplementary Table 1. Calculation of MET-min/wk variables for IPAQ to calculate a continuous score

Category	Weighted Met level	Calculation of MET-minutes/week
Walking	3.3 METs	3.3 x walking minutes/week
Moderate	4.0 METs	4.0 x moderate minutes/week
Vigorous	8.0 METS	8.0 x vigorous minutes/week
TOTAL MET-min/wk		Walking MET-min/wk + Moderate MET-min/wk + Vigorous MET-min/wk

Supplementary Figure 1. Elbow plot of Bayesian Information Criteria and other indices for



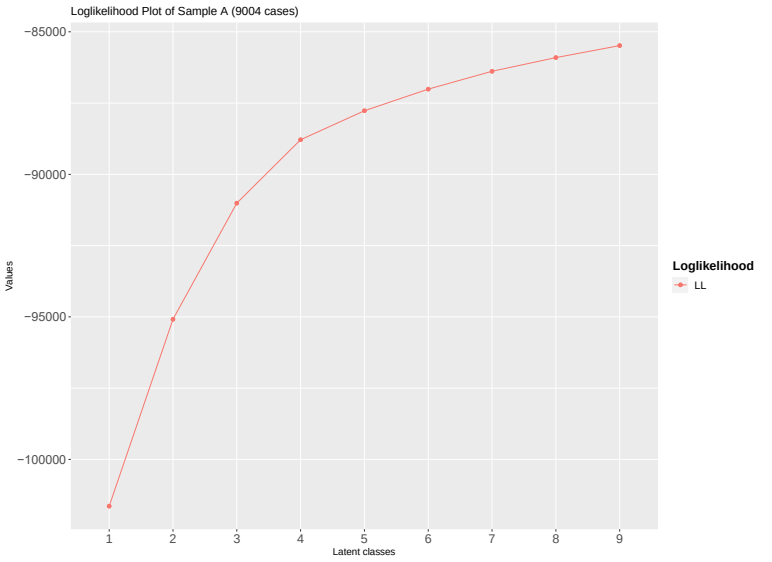
model fitting

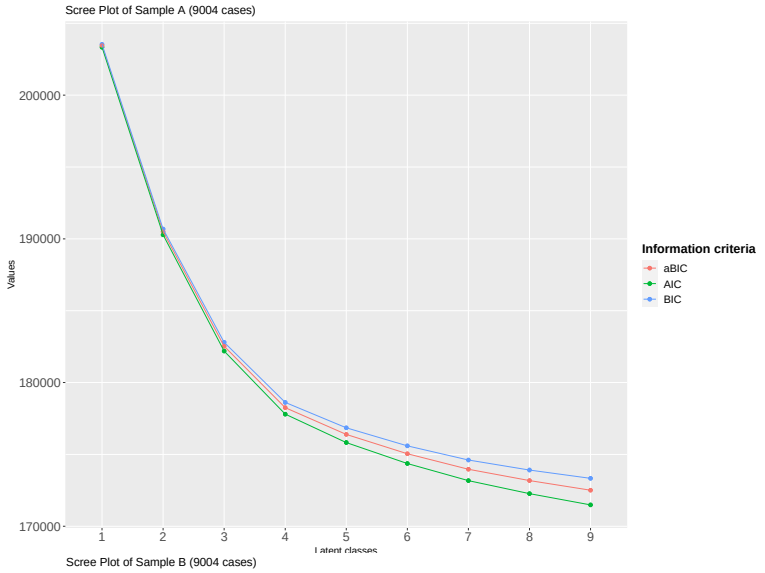
Note. BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion

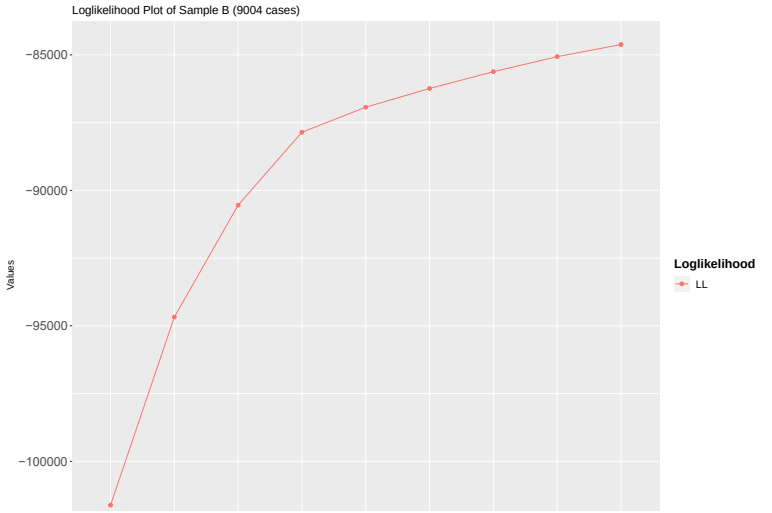
Supplementary Section A: Split-half replication

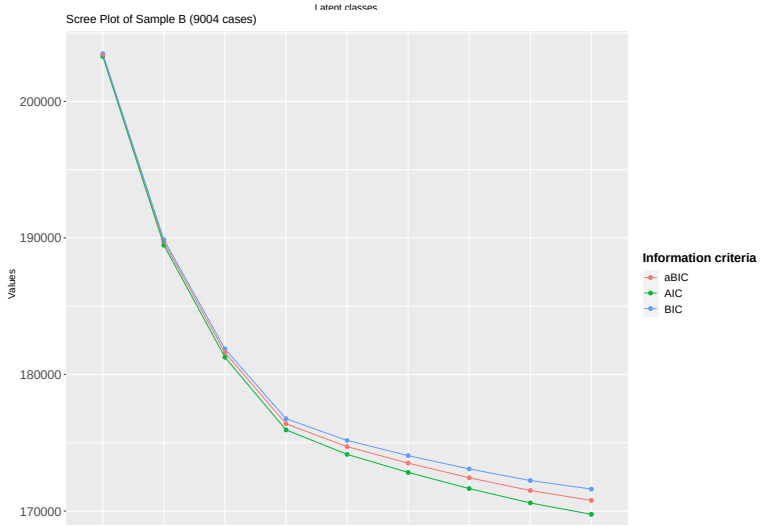
To ensure the reliability of our class solution, the sample (n=18008) was randomly split into halves and separate, sequential latent class models were conducted on both random subsamples. These subsamples were named Sample A (n=9004) and Sample B (n=9004). We then compared the model fit statistics (such as the Bayesian information criterion, sample size adjusted BIC, consistent Akaike information criterion, and loglikelihoods) of the three samples using screeplots. If the final latent class solution (i.e. a seven-class model) were consistently identified as the best fit in all three samples, this would provide evidence for the replicability of the latent classes identified in the analysis. Fig S2 shows the screeplots for the model fit statistics in the three samples. Screeplots of model fit statistics for one to nine class solutions in the random subsamples and those for the whole sample are shown in Fig S2. In all three samples, the decline in the information criteria levelled off around five classes, providing initial evidence for replication of the results across the three samples. Additionally, we compared the structures (i.e. the prevalence of latent classes and the probabilities of item responses) of the five-class model with the six-class and seven-class models in the three samples (see Supplementary Figures 3-5). We found that both the five- and seven-class models replicated with maximum stability (i.e. had a similar latent class structure) across all three samples. On the other hand, the six-class model was unstable as it did not replicate across all the three samples.

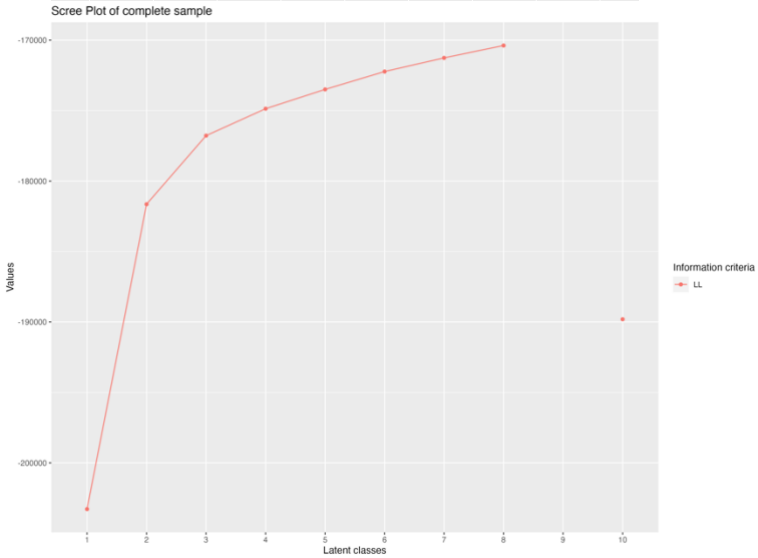
Supplementary Figure 2. Screeplots of Bayesian information criterion (BIC), Sample size adjusted BIC (aBIC); Consistent Akaike information criterion (AIC) and loglikelihoods for i) Sample A (n=9004, 50% random split sample), ii) Sample B (n=9004, 50% random split sample) and iii) Complete Sample (n=18008).

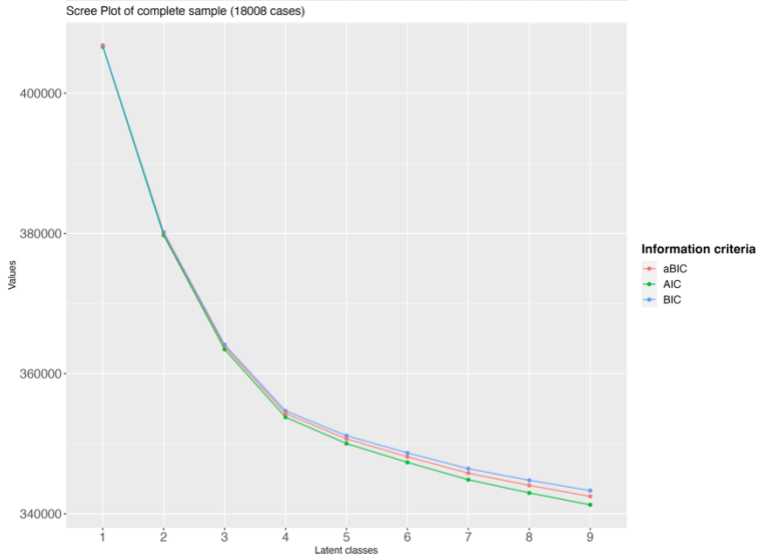




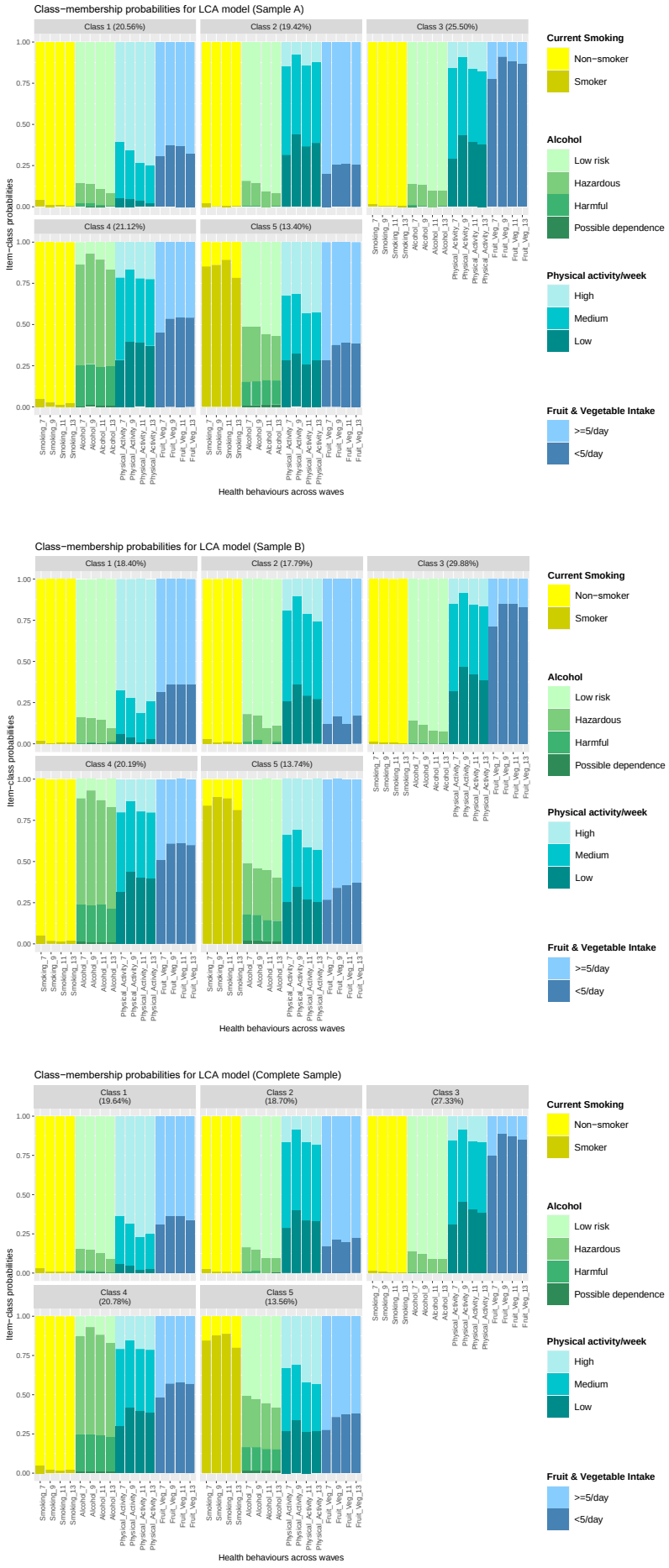




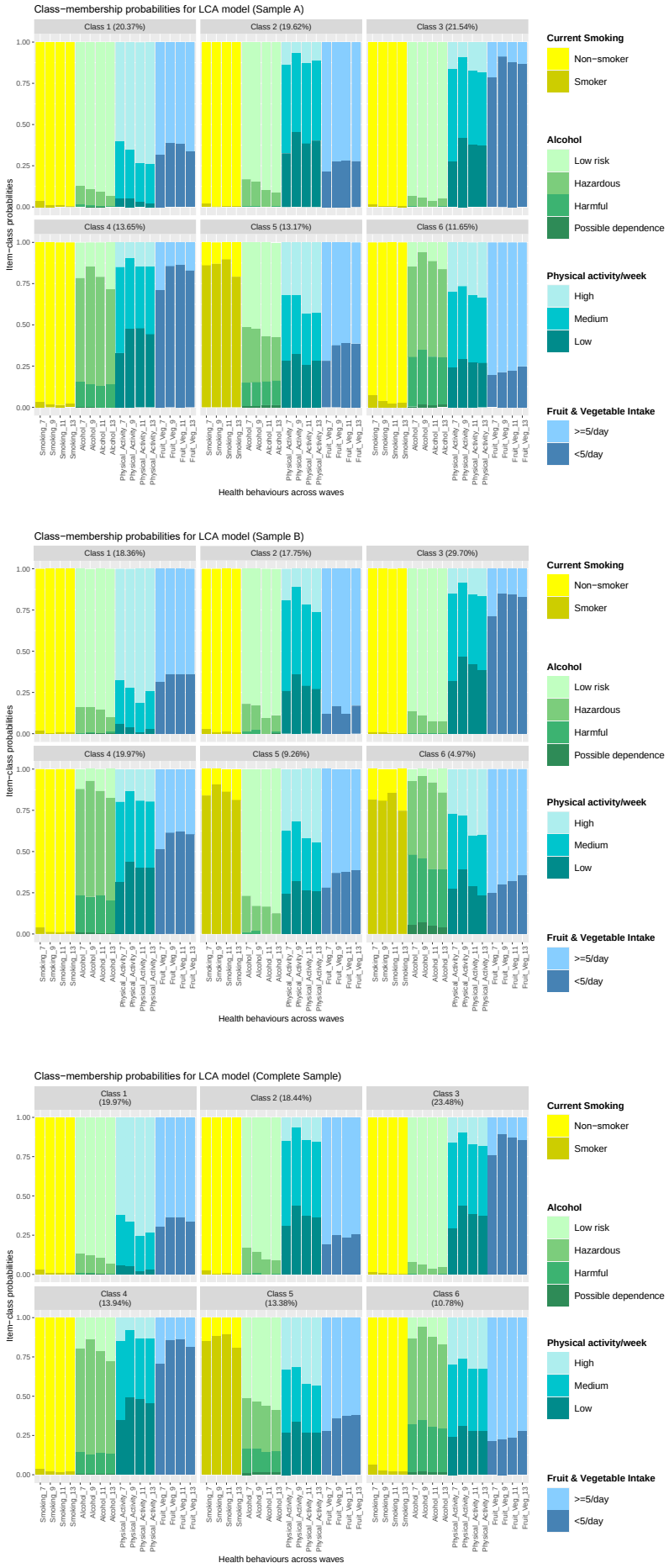




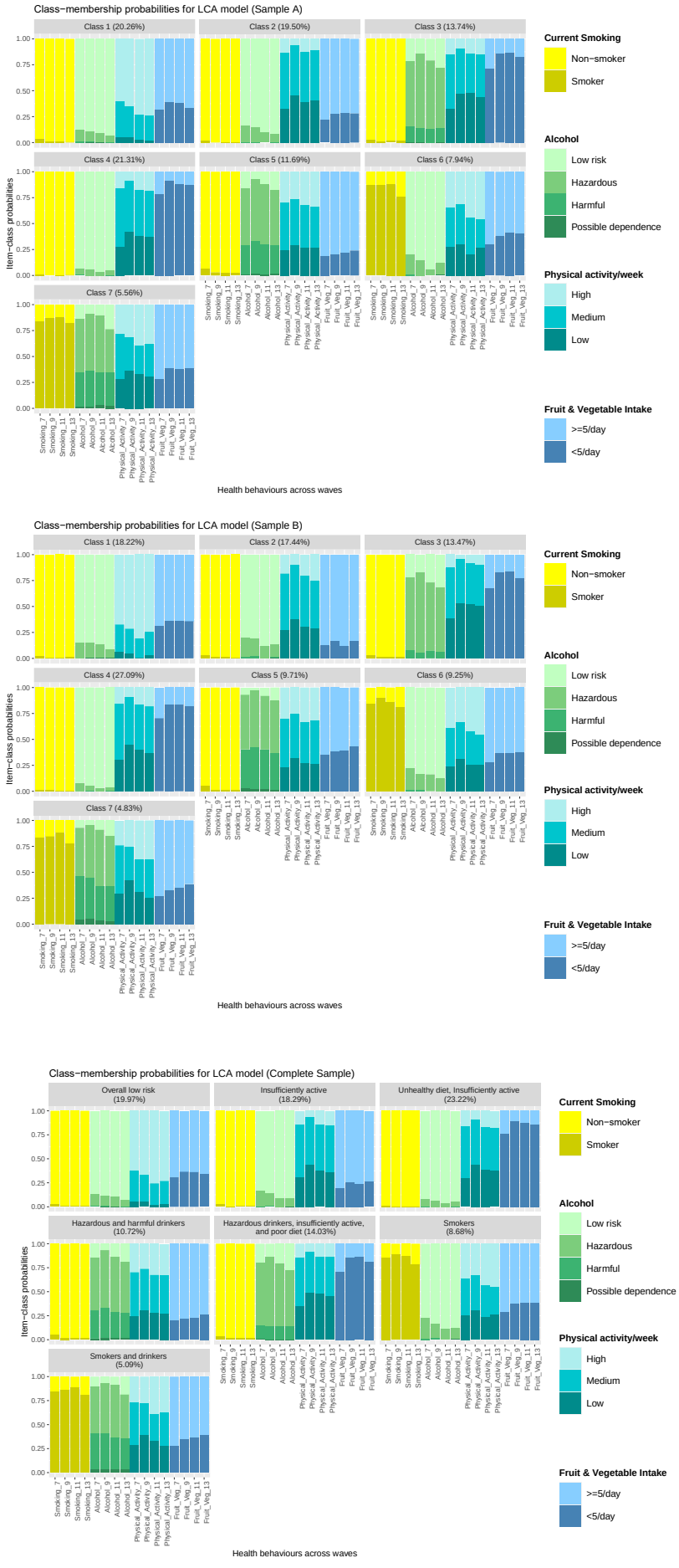
Supplementary Figure 3. 5-class model solutions for Sample A (n=9004, 50% random split sample), Sample B (n=9004, 50% random split sample) and Complete Sample (n=18008)



Supplementary Figure 4. 6-class model solutions for Sample A (n=9004, 50% random split sample), Sample B (n=9004, 50% random split sample) and Complete Sample (n=18008)



Supplementary Figure 5. 7-class model solutions for Sample A (n=9004, 50% random split sample), Sample B (n=9004, 50% random split sample) and Complete Sample (n=18008)



Supplementary Table 2. Prevalence of chronic conditions at baseline Wave 7 and Wave 13

Baseline socio-demographic characteristics		**Sample (n=18008)**
Disease status at Wave 7		N (%)
Multimorbidity (>= 2 chronic conditions)	Has condition	2207 (12.3)
	Missing	3519 (19.5)
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)	Has condition	2627 (14.6)
	Missing	4389 (24.4)
Cancer (All types)	Has condition	277 (1.5)
	Missing	4640 (25.8)
Obesity-related (Diabetes, High Blood pressure)	Has condition	3106 (17.2)
	Missing	4193 (233)
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)	Has condition	1701 (9.4)
	Missing	4590 (25.5)
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)	Has condition	843 (4.7)
	Missing	4618 (25.6)
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)	Has condition	908 (5)
	Missing	4618 (25.6)
Disease status at Wave 13		
Multimorbidity (>= 2 chronic conditions)	Has condition	613 (3.4)
	Missing	169 (0.9)
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)	Has condition	601 (3.3)
	Missing	169 (0.9)
Cancer (All types)	Has condition	217 (1.2)
	Missing	169 (0.9)
Obesity-related (Diabetes, High Blood pressure)	Has condition	1124 (6.2)
	Missing	169 (0.9)
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)	Has condition	448 (2.5)
	Missing	169 (0.9)
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)	Has condition	321 (1.8)
	Missing	169 (0.9)
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)	Has condition	726 (4)
	Missing	169 (0.9)
Morbid Obesity (BMI ≥40)	Has condition	640(3.4)
	Missing	103(0.4)

Supplementary Section B: Sub-sample aged 49 years or younger

Health behaviour clusters and within-cluster changes

After fitting models with one to nine latent classes, a model with four classes was considered the best fit (model fit statistics are shown in Supplementary Table 3). The VLMR-LMR test supported a four-class solution. However, the values of the BIC, SABIC, and CAIC continued to decrease as the number of classes increased, suggesting improved model fit (Chen et al., 2017). As the elbow-point in information criteria occurred around four classes (screeplot shown in Supplementary Fig S2), we opted for the four-class model. The entropy and smallest average latent class posterior probability fell within the recommended range (value ≥0.8) for the four-class model (Nylund-Gibson and Choi, 2018). Furthermore, sensitivity analysis using split-halves of the sample also indicated that the 4-class solution was the optimal solution, as identical classes were uncovered in each split-half. The four distinct health-risk behaviour classes—hereafter referred to as clusters—were assigned labels. Supplementary Figure 6 depicts these clusters.

Participants in the *Overall low-risk* cluster (29.3% of the sample) maintained adequate (≥5 per day) fruit and vegetable intake and medium-high levels of physical activity across time. They also had a consistently high probability of drinking at low-risk levels, and the likelihood of consuming alcohol at hazardous levels fell across waves.

The largest proportion of participants fell in the *Insufficiently active, Poor diet* cluster (32.2%). Individuals in this cluster displayed low-medium levels of physical activity but did not consume the recommended portions of fruit and vegetables. They had low-risk alcohol intake and the probability of consuming alcohol at hazardous levels fell across waves.

Participants in the *Insufficiently active, Hazardous drinkers* cluster (22.6%) had a high likelihood of consuming alcohol at hazardous or harmful levels and displayed low-medium levels of physical activity.

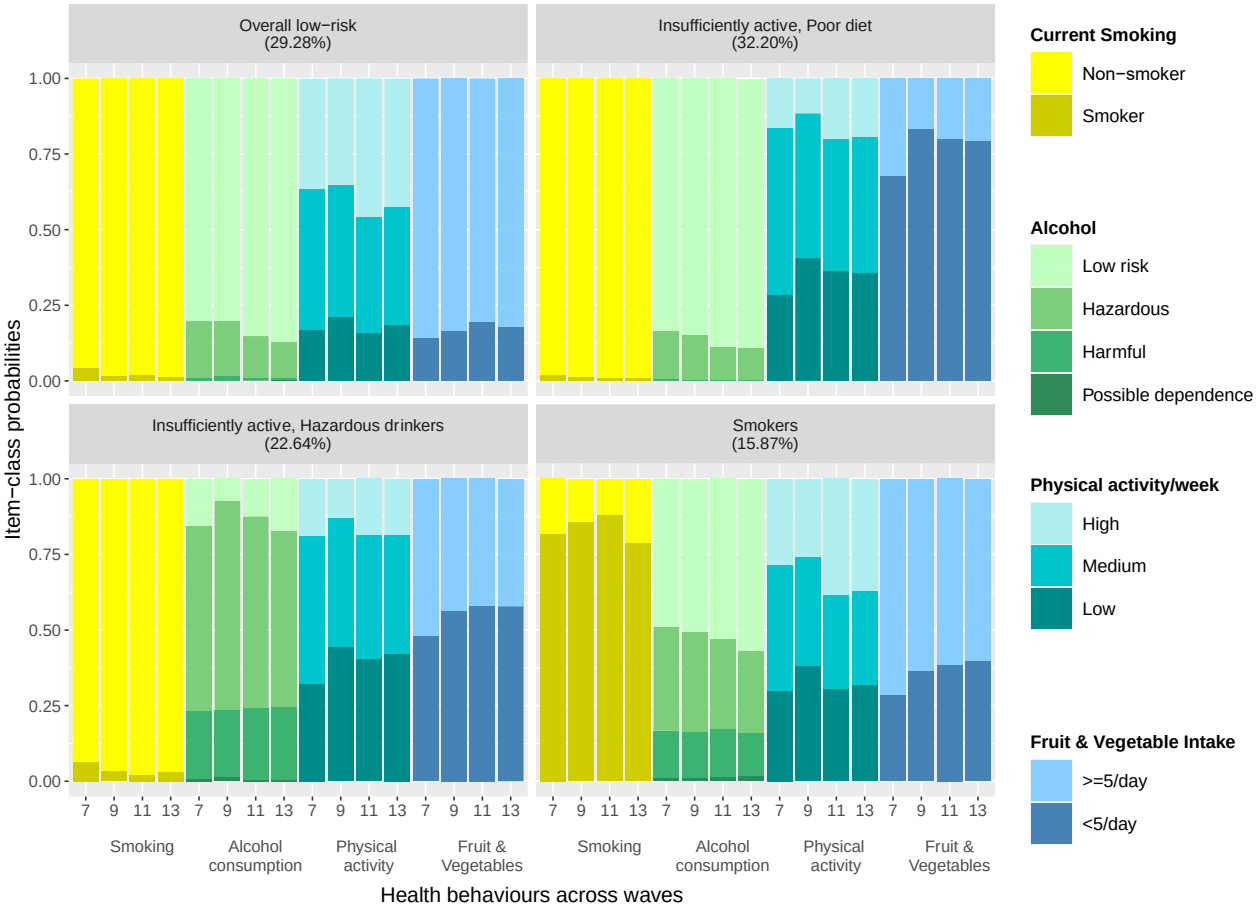
Finally, participants characterised as *Smokers* (15.9%) had a high probability of smoking, increasing over time then declining over the last wave and adequate fruit and vegetable intake. They had the same probability of engaging in physical activity across all three levels (i.e. high, medium, low). Notably, participants in this cluster exhibited a 50% probability of engaging in alcohol consumption at low-risk levels, and the around 50% probability of consuming alcohol at hazardous to harmful levels.

Supplementary Table 3. Model fit evaluation information for choosing a latent class model

K	LL	CAIC	BIC	SABIC	AWE	VLMR LRT p-value	Entropy	Smallest average latent class posterior probability	Smallest class size (%)
1	-94219.8	188496	188692	188603	188591.033	-	-	1	100
2	-88255.3	176625	177024	176843	176818.97	<0.001	0.957	0.975	17.6
3	-84983.9	170140	170742	170469	170432.90	<0.001	0.856	0.907	16.5
4	-82941.1	166112	166917	166552	166504.28	<0.001	0.821	0.881	15.9
5	-82260.8	164810	165818	165360	165300.44	0.062	0.798	0.829	12.6
6	-81604.7	163555	164767	164217	164145.16	0.114	0.804	0.838	6.0
7	-80963.9	162332	163746	163104	163020.50	0.078	0.787	0.792	6.0
8	-80606.7	161675	163293	162559	162462.80	0.233	0.78	0.778	6.0
9	-80285.1	161090	162910	162084	161976.55	0.080	0.777	0.755	5.1

Note: n=8110; K = number of classes (the nine-class model failed to converge); LL = model log-likelihood; BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; p-value significance <0.05.

Supplementary Figure 6. Four-class model reflecting different clusters of health behaviour across time



Clusters and socio-demographic characteristics (Sub-sample aged 49 years or younger)

Supplementary Table 4 displays the socio-demographic composition of participants in each of the identified clusters and the results from the adjusted multinomial logistic regressions that examined associations between each socio-demographic characteristic and cluster membership.

Compared to participants in the *Overall low-risk* cluster (which served as the reference group), participants in all other clusters were more likely to be older.

Women made up around two-thirds of the *Insufficiently Active, Poor diet* cluster. while men comprised two-thirds of the *Hazardous and harmful drinkers.* Compared to the *Overall low-risk* cluster (reference group), participants characterised as *Insufficiently Active, Hazardous drinkers* were more likely to be male.

Compared to the *Overall low-risk* cluster, individuals in *Insufficiently Active, Poor diet* cluster were two times more likely to belong to other White ethnicities (i.e. other than White British). In contrast, participants characterised *as Insufficiently Active, Hazardous drinkers* and *Smokers* were more likely to be White British than belong to the other ethnicities studied (i.e. Indian, Pakistani or Bangladeshi, Black, or Other).

Participants characterised as *Smokers* were 2-3 times more likely to be educated to levels other than degree level, compared to the *Overall low-risk* cluster, and instead had the smallest proportion of individuals educated to degree levels and higher (21.8%), which was less than half of individuals in other clusters. In contrast, Participants characterised as *Insufficiently Active, Poor diet* were less likely to have to have their highest educational qualification be O-levels or no qualification, compared the *Overall low-risk cluster*.

Finally, participants characterised as *Insufficiently Active, Hazardous drinkers* were more likely to belong the highest quintile of household income than any other quintile, compared to the *Overall low-risk cluster*. On the other hand, individuals in the *Smokers* cluster were 2-3 times more likely to fall in the 2nd,3rd or 4th quintile compared to the highest quintile.

Supplementary Table 4. Demographics and Odds ratios from Multinomial Logistic regressions examining the association between socio-demographic predictors and cluster membership.

	Overall low-risk (Ref)	Insufficiently Active, Poor diet		Insufficiently Active, Hazardous drinkers		Smokers	
	29.28%	32.21%		22.64%		15.87%	
			OR [95% C.I.]		OR [95% C.I.]		OR [95% C.I.]
Age - Mean (s.d.)	31.90 (30.44)	34.25 (28.73)	**1.02 [1.01, 1.03]**	34.24 (35.03)	**1.02 [1.01, 1.03]**	33.28 (41.43)	**1.02 [1.01, 1.03]**
Sex							
Male (Ref)	44.9%	36.3%		60.1%		46.0%	
Female	55.1%	63.7%	**1.44 [1.18, 1.76]**	39.9%	**0.54 [0.06, 0.66]**	54.0%	0.89 [0.7, 1.12]
Ethnicity							
White - British (Ref)	81.2%	83.5%		91.6%		87.0%	
White - Other	3.1%	6.6%	**2.09 [1.25, 3.51]**	3.3%	0.97 [0.48, 1.95]	4.3%	1.21 [0.67, 2.19]
Indian	3.8%	3.3%	0.87 [0.55, 1.37]	1.0%	**0.22 [0.11, 0.44]**	0.9%	**0.26 [0.13, 0.51]**
Pakistani or Bangladeshi	4.4%	1.1%	**0.30 [0.18, 0.51]**	1.1%	**0.25 [0.12, 0.52]**	1.9%	**0.32 [0.19, 0.55]**
Black - African/Caribbean	2.9%	1.8%	0.70 [0.39, 1.25]	0.7%	**0.23 [0.07, 0.81]**	1.3%	**0.32 [0.16, 0.67]**
Other	4.6%	3.7%	0.76 [0.48, 1.20]	2.3%	**0.45 [0.25, 0.81]**	4.6%	1.07 [0.59, 1.92]
Educational Qualification							
Degree level or higher (Ref)	41.0%	52.9%		46.4%		21.8%	
A-levels or equivalent	27.7%	26.5%	0.91 [0.71, 1.15]	28.2%	1.10 [0.85, 1.44]	29.4%	**1.91 [1.39, 2.61]**
O-levels or equivalent	22.4%	15.6%	**0.66 [0.5, 0.87]**	19.6%	0.97 [0.73, 1.29]	32.8%	**2.35 [1.72, 3.21]**
Others	3.9%	3.1%	0.68 [0.39, 1.20]	3.6%	0.91 [0.49, 1.71]	10.4%	**3.78 [2.20, 6.50]**
None	5.0%	1.9%	**0.38 [0.2, 0.74]**	2.2%	0.57 [0.28, 1.17]	5.6%	1.75 [0.98, 3.14]
Household Income							
Highest Quintile (Ref)	13.3%	18.9%		22.7%		5.6%	
4th Quintile	18.0%	22.2%	**0.55 [0.39, 0.78]**	21.6%	**0.48 [0.33, 0.69]**	10.7%	**3.13 [1.95, 5.04]**
3rd Quintile	20.7%	22.7%	0.74 [0.54, 1.02]	22.1%	**0.47 [0.33, 0.66]**	19.7%	**2.18 [1.36, 3.49]**
2nd Quintile	24.5%	21.8%	0.83 [0.61, 1.14]	18.0%	**0.62 [0.44, 0.87]**	26.7%	**1.91 [1.17, 3.11]**
Lowest Quintile	23.5%	14.4%	0.89 [0.65, 1.22]	15.6%	**0.67 [0.49, 0.93]**	37.3%	1.27 [0.76, 2.11]

Note: **Bold values** indicate statistically significant values at p=0.05

Health behaviour clusters and disease status (Sub-sample aged 49 years or younger)

After controlling for both sociodemographic characteristics and baselines, only the prevalence of respiratory disorders differed significantly across clusters — omnibus Wald test χ^2(df = 3) = 12.957, p=0.005. The prevalence of morbid obesity also differed across clusters after controlling for sociodemographic characteristics, omnibus Wald test χ^2(df = 3) = 9.444, p=0.024. To further investigate which clusters differed in the proportions of participants with each disease profile, we conducted pairwise comparisons for results adjusted for disease status and socio-demographic variables between clusters for respiratory disorders as well as for morbid obesity, as shown in Supplementary Table 5.

Participants characterised as *Smokers* had a higher prevalence of respiratory disorders than the *Insufficiently Active, Hazardous drinkers* and the *Overall low-risk cluster*. Participants characterised as *Overall low-risk* had a higher prevalence of morbid obesity than the *Insufficiently Active, Hazardous drinkers*.

Supplementary Table 5. Pairwise comparisons of disease status prevalence across clusters

	Overall low-risk	Insufficiently Active, Poor diet	Insufficiently Active, Hazardous drinkers	Smokers	Wald's omnibus test χ^2(df = 3)	Wald's omnibus test (p value)	Sample size
	29.28%	32.21%	22.64%	15.87%			
Multimorbidity (>= 2 chronic conditions)							
Unadjusted	0.014	0.01	0.006	0.018	4.053	p=0.256	n=8110
Adjusted for socio-demographic covariates	0.013	0.011	0.006	0.017	1.463	p=0.691	n=8110
*Adjusted for socio-demographic covariates and baseline disease	0.01	0.014	0.008	0.018	1.744	p=0.627	n=6297
Cancers (All types)							
Unadjusted	0.003	0.006	0.001	0.004	2.581	p=0.461	n=8110
Adjusted for socio-demographic covariates	0.003	0.006	0.001	0.004	2.8	p=0.424	n=8110
*Adjusted for socio-demographic covariates and baseline disease	0.002	0.005	0.001	0.006	3.747	p=0.290	n=6043
Obesity-related (Diabetes, High Blood pressure)							
Unadjusted	0.024	0.028	0.019	0.027	1.784	p=0.618	n=8110
Adjusted for socio-demographic covariates	0.024	0.028	0.019	0.027	1.874	p=0.599	n=8110
*Adjusted for socio-demographic covariates and baseline disease	0.024	0.029	0.023	0.022	3.76	p=0.289	n=6133
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)							
Unadjusted	0.017	0.021	0.014	0.052	18.655	**p<0.001**	n=8110
Adjusted for socio-demographic covariates	0.017	0.021	0.015	0.051	11.952	**p=0.008**	n=8110
*Adjusted for socio-demographic covariates and baseline disease	0.018 [a]	0.023	0.015 [b]	0.057 [a,b]	12.957	**p=0.005**	n=6092
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)							
Unadjusted	0.011	0.007	0.007	0.02	8.641	**p=0.035**	n=8110
Adjusted for socio-demographic covariates	0.011	0.007	0.007	0.019	5.782	p=0.123	n=8110
*Adjusted for socio-demographic covariates and baseline disease	0.012	0.007	0.008	0.024	5.258	p=0.154	n=6089
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)							
Unadjusted	0.005	0.001	0.003	0.009	5.32	p=0.150	n=8110
Adjusted for socio-demographic covariates	0.005	0.001	0.003	0.009	3.343	p=0.342	n=8110
*Adjusted for socio-demographic covariates and baseline disease	0.003	0.002	0.003	0.009	1.949	p=0.583	n=6040
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)							
Unadjusted	0.037	0.032	0.014	0.038	9.742	**p=0.021**	n=8110
Adjusted for socio-demographic covariates	0.037	0.031	0.015	0.038	7.379	p=0.061	n=8110
*Adjusted for socio-demographic covariates and baseline disease	0.033	0.032	0.018	0.038	2.628	p=0.453	n=6070
Morbid Obesity (BMI ≥ 40)							
Unadjusted	0.053 [a]	0.036	0.022 [a]	0.046	11.484	**p=0.009**	n=8110

Adjusted for socio-demographic covariates	0.053 [a]	0.035	0.023 [a]	0.045	9.444	**p=0.024**	n=8110

Note: Socio-demographic covariates adjusted for include: sex, age, own occupation, education level and monthly household income quintile.
[a,b] indicates statistically significant pairwise differences in disease prevalence at Bonferroni corrected significance level of p=0.008.
Bold values indicate a statistically significant omnibus wald's test at significance level of p=0.05.

Supplementary Section C: Sub-sample aged 50 years or older

Health behaviour clusters and within-cluster changes

After fitting models with one to nine latent classes, a model with five classes was considered the best fit (model fit statistics are shown in Supplementary Table 6). The VLMR-LMR test supported a five-class solution. However, the values of the BIC, SABIC, and CAIC continued to decrease as the number of classes increased, suggesting improved model fit (Chen et al., 2017). As the elbow-point in information criteria occurred around four and five classes, so we opted for the five-class model. The entropy and smallest average latent class posterior probability fell within the recommended range (value ≥0.8) for the five-class model (Nylund-Gibson and Choi, 2018). Furthermore, sensitivity analysis using split-halves of the sample also indicated that the 5-class solution was the optimal solution, as identical classes were uncovered in each split-half. The five distinct health-risk behaviour classes—hereafter referred to as clusters—were assigned labels. Supplementary Figure 7 depicts these clusters.

Supplementary Table 6. Model fit evaluation information for choosing a latent class model

K	LL	CAIC	BIC	SABIC	AWE	VLMR LRT p-value	Entropy	Smallest average latent class posterior probability	Smallest class size (%)
1	-107841.3	215739	215940	215851	215836.455	-	-	1	100
2	-100271.5	200657	201068	200886	200856.286	<0.001	0.988	0.992	11.7
3	-95334.7	190841	191461	191187	191142.017	<0.001	0.895	0.931	11.4
4	-92170.1	184570	185398	185033	184972.248	<0.001	0.837	0.885	11.2
5	-90948.5	182185	183222	182764	182688.379	<0.001	0.821	0.805	11.2
6	-90249.0	180844	182090	181540	181448.63	0.342	0.814	0.802	6.5
7	-89600.7	179605	181060	180418	180311.401	0.053	0.797	0.792	5.1
8	-88962.7	178387	180051	179317	179194.811	0.006	0.801	0.793	4.0
9	-88442.2	177404	179276	178450	178313.262	0.482	0.795	0.773	4.0

Note: n=9898; K = number of classes (the nine-class model failed to converge); LL = model log likelihood; BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; p-value significance <0.05.

Participants in the *Overall low-risk* cluster (22.9% of the sample) maintained adequate (≥5 per day) fruit and vegetable intake and high levels of physical activity across time. They also had a consistently high probability of drinking at low-risk levels, and the proportion of people consuming alcohol at hazardous levels fell across waves. Notably, the proportion of individuals engaging in high levels of physical activity rose over time.

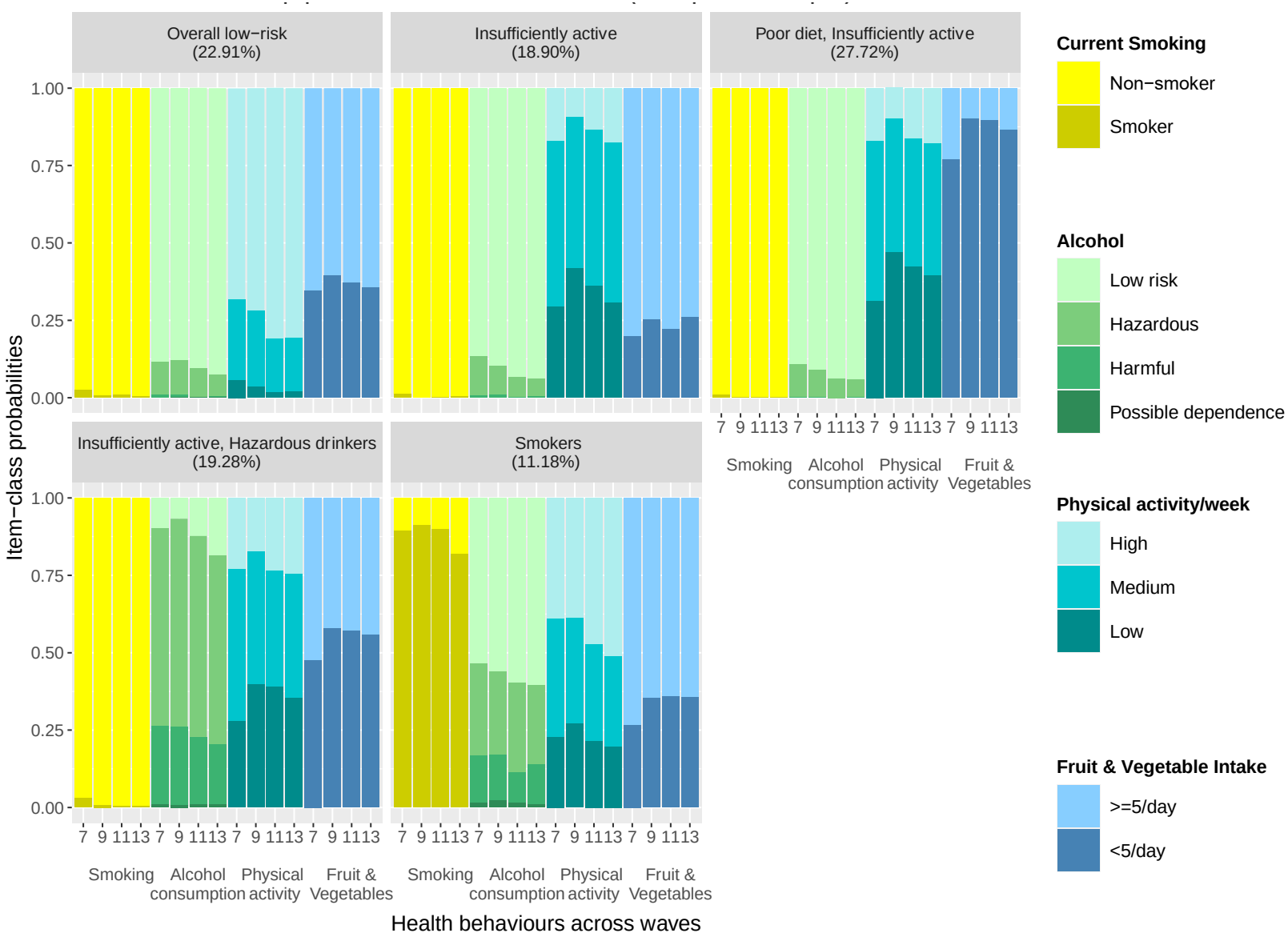
Participants in the *Insufficiently active* cluster (18.9%) had a high likelihood of displaying low-medium levels of physical activity, but no other engagement in risky behaviour.

The largest proportion of participants fell in the *Insufficiently active, Poor diet* cluster (27.7%). Individuals in this cluster displayed low-medium levels of physical activity and did not consume the recommended portions of fruit and vegetables. They had low-risk alcohol intake and the proportion of people consuming alcohol at hazardous levels fell across waves.

Participants in the *Insufficiently active, Hazardous drinkers* cluster (19.3%) had a consistently high probability of consuming alcohol at hazardous levels and a high likelihood of displaying low-medium levels of physical activity.

Finally, participants characterised as *Smokers* (11.2%) had a high probability of smoking. They had the same probability of engaging in physical activity across at high-medium levels. Notably, participants in this cluster maintained adequate (≥5 per day) fruit and vegetable intake.

Supplementary Figure 7. Five-class model reflecting different clusters of health behaviour across time



Clusters and socio-demographic characteristics (Sub-sample aged 50 years or older)

Supplementary Table 7 displays the socio-demographic composition of participants in each of the identified clusters.

Compared to participants in the *Overall low-risk* cluster (which served as the reference group), participants in all other clusters were more likely to be younger.

Women made up around two-thirds the *Overall low risk*. while men comprised two-thirds of the *Hazardous drinkers and Insufficiently active* cluster. Compared to the *Overall low-risk* cluster (reference group), participants in all clusters were less likely be female.

Compared to the *Overall low-risk* cluster, individuals in *Insufficiently Active* cluster and *Hazardous drinkers and Insufficiently active* were less likely to belong to non-White ethnicities (i.e. Indian, Pakistani or Bangladeshi, Black, or Other). Similarly, participants characterised as Smokers and *Poor diet and Insufficiently active* were less likely to be Indian, compared to the *Overall low-risk* cluster.

Participants characterised as *Smokers* were 1.5-2 times more likely to be educated to levels other than degree level, compared to the *Overall low-risk* cluster, and instead had the smallest proportion of individuals educated to degree levels and higher (21.8%), which was less than half of individuals in other clusters. In contrast, participants characterised as *Insufficiently Active, Poor diet and Insufficiently Active*, and *Hazardous drinkers and Insufficiently active* were less likely to have to have their highest educational qualification be O-levels, others or no qualification, compared the *Overall low-risk cluster*.

Finally, participants characterised as participants characterised as *Insufficiently Active, Poor diet and Insufficiently Active*, and *Hazardous drinkers and Insufficiently active* were more likely to belong the highest quintile of household income than any other quintile, compared to the *Overall low-risk cluster*.

Supplementary Table 7. Demographics and Odds ratios from Multinomial Logistic regressions examining the association between socio-demographic predictors and cluster membership.

	Overall Low-risk	Insufficiently Active			Poor diet and Insufficiently active			Hazardous drinkers and insufficiently active			Smokers		
	22.91%	18.90%			27.72%			19.28%			11.19%		
			OR	[95% C.I]		OR	[95% C.I]		OR	[95% C.I]		OR	[95% C.I]
Age (Mean)	69.059	63.62	0.95	[0.94,0.96]	63.22	0.95	[0.94,0.97]	60.33	0.91	[0.90,0.92]	60.56	0.9	[0.88,0.91]
Sex													
Male (Ref)	31.5%	54.5%	*Ref*		39.3%	*Ref*		67.7%	*Ref*		46.1%	*Ref*	
Female	68.5%	45.5%	0.37	[0.30,0.47]	60.7%	0.73	[0.60,0.89]	32.3%	0.21	[0.17,0.26]	53.9%	0.49	[0.39,0.62]
Ethnicity													
White - British (Ref)	93.6%	96.0%	*Ref*		94.5%	*Ref*		97.1%	*Ref*		94.7%	*Ref*	
White - Other	2.1%	2.9%	1.02	[0.57,1.82]	3.3%	1.09	[0.68,1.75]	2.2%	0.71	[0.40,1.26]	2.9%	1.03	[0.58,1.83]
Indian	1.8%	0.7%	0.26	[0.11,0.62]	0.8%	0.31	[0.17,0.57]	0.3%	0.08	[0.02,0.31]	0.6%	0.23	[0.08,0.64]
Pakistani or Bangladeshi	0.9%	0.1%	0.12	[0.02,0.61]	0.2%	0.36	[0.07,1.78]	0.1%	0.15	[0.03,0.93]	0.4%	0.28	[0.10,0.77]
Black - African/Caribbean	1.6%	0.3%	0.17	[0.06,0.47]	1.2%	0.56	[0.30,1.07]	0.3%	0.12	[0.04,0.36]	1.4%	0.67	[0.34,1.34]
Other	1.5%	0.7%	0.32	[0.13,0.76]	1.6%	0.63	[0.34,1.20]	0.5%	0.15	[0.05,0.42]	1.9%	0.84	[0.42,1.70]
Educational Qualification													
Degree level or higher (Ref)	19.8%	35.6%	*Ref*		46.3%	*Ref*		43.8%	*Ref*		16.5%	*Ref*	
A-levels or equivalent	10.8%	16.2%	0.82	[0.60,1.13]	16.6%	0.75	[0.57,0.98]	18.7%	0.9	[0.67,1.21]	14.2%	1.53	[1.07,2.18]
O-levels or equivalent	19.7%	21.8%	0.71	[0.54,0.94]	16.5%	0.43	[0.34,0.55]	19.7%	0.65	[0.50,0.85]	25.4%	1.63	[1.20,2.23]
Others	18.2%	15.3%	0.64	[0.46,0.90]	12.5%	0.44	[0.33,0.58]	9.7%	0.48	[0.34,0.67]	16.8%	1.54	[1.07,2.20]
None	31.5%	11.1%	0.35	[0.25,0.50]	8.1%	0.21	[0.16,0.29]	8.1%	0.39	[0.28,0.56]	27.1%	2.15	[1.53,3.04]
Household Income													
Highest Quintile (Ref)	8.5%	19.0%	*Ref*		23.6%	*Ref*		32.0%	*Ref*		10.9%	*Ref*	
4th Quintile	12.7%	19.0%	0.65	[0.46,0.91]	23.0%	0.4	[0.30,0.53]	23.4%	0.23	[0.16,0.32]	12.6%	1.04	[0.72,1.51]
3rd Quintile	21.2%	18.0%	0.51	[0.36,0.72]	19.0%	0.42	[0.31,0.56]	16.7%	0.25	[0.18,0.34]	16.8%	0.81	[0.56,1.17]

2nd Quintile	28.4%	20.2%	0.49	[0.35,0.69]	18.1%	0.45	[0.34,0.60]	15.4%	0.29	[0.21,0.39]	26.3%	0.64	[0.44,0.93]
Lowest Quintile	29.2%	23.8%	0.75	[0.53,1.06]	16.3%	0.76	[0.56,1.02]	12.5%	0.56	[0.41,0.76]	33.4%	0.76	[0.51,1.13]

Note: Bold values indicate statistically significant values at p=0.05

Health behaviour clusters and disease status (Sub-sample aged 50 years or older)

After controlling for both sociodemographic characteristics and baselines, only the prevalence of multimorbidity differed significantly across clusters — omnibus Wald test $\chi^2(df = 4) = 13.895$, $p=0.008$. The prevalence of morbid obesity also differed across clusters after controlling for sociodemographic characteristics, omnibus Wald test $\chi^2(df = 4) = 78.263$, $p<0.001$. To further investigate which clusters differed in the proportions of participants with multimorbidity, we conducted pairwise comparisons for results adjusted for disease status and socio-demographic variables, as shown in Supplementary Table 8.

Participants characterised as *the Overall low-risk cluster* had a higher prevalence of multimorbidity than the *Insufficiently Active* and *Poor diet and insufficiently active*. Participants characterised as *Overall low-risk cluster* had a higher prevalence of morbid obesity compared to all other clusters. Smokers also had a higher prevalence of morbid obesity than participants characterised as *Hazardous drinkers and insufficiently active.*

Supplementary Table 8. Pairwise comparisons of disease status prevalence across clusters

	Overall low-risk	Insufficiently Active	Poor diet and insufficiently active	Hazardous drinkers and insufficiently active	Smokers	Wald's test $\chi2(df = 4)$	Wald's test p-value	Sample size
	22.91%	18.90%	27.72%	19.28%	11.19%			
Multimorbidity (>= 2 chronic conditions)								
Unadjusted	0.088	0.039	0.038	0.031	0.071	50.82	**p<0.001**	n=9898
Adjusted for socio-demographic covariates	0.085	0.04	0.039	0.032	0.071	20.663	**p<0.001**	n=9898
Adjusted for socio-demographic covariates and baseline disease	0.09[a,b]	0.043[a]	0.041[b]	0.032	0.076	13.895	**p=0.008**	n=8192
Cancers (All types)								
Unadjusted	0.023	0.018	0.016	0.017	0.025	3.664	p=0.453	n=9898
Adjusted for socio-demographic covariates	0.022	0.018	0.016	0.018	0.025	3.953	p=0.413	n=9898
Adjusted for socio-demographic covariates and baseline disease	0.023	0.02	0.017	0.02	0.029	4.626	p= 0.328	n=7325
Obesity-related (Diabetes, High Blood pressure)								
Unadjusted	0.116	0.075	0.081	0.079	0.099	14.7	**p=0.005**	n=9898
Adjusted for socio-demographic covariates	0.115	0.074	0.082	0.079	0.099	8.062	p=0.089	n=9898
Adjusted for socio-demographic covariates and baseline disease	0.119	0.082	0.086	0.081	0.104	2.848	p=0.584	n=7682
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)								
Unadjusted	0.04	0.015	0.021	0.017	0.061	35.495	**p<0.001**	n=9898
Adjusted for socio-demographic covariates	0.04	0.015	0.021	0.017	0.061	22.896	**<0.001**	n=9898
Adjusted for socio-demographic covariates and baseline disease	0.044	0.019	0.023	0.016	0.055	8.082	p=0.089	n=7326
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)								
Unadjusted	0.08	0.028	0.043	0.044	0.064	30.255	**p<0.001**	n=9898
Adjusted for socio-demographic covariates	0.078	0.03	0.043	0.046	0.064	17.016	**p=0.002**	n=9898

Adjusted for socio-demographic covariates and baseline disease	0.081	0.031	0.049	0.05	0.067	8.019	p=0.091	n=7530
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)								
Unadjusted	0.057	0.027	0.02	0.024	0.046	29.324	**p<0.001**	n=9898
Adjusted for socio-demographic covariates	0.055	0.028	0.022	0.025	0.046	11.866	**p=0.018**	n=9898
Adjusted for socio-demographic covariates and baseline disease	0.055	0.034	0.022	0.024	0.045	4.909	p=0.297	n=7350
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)								
Unadjusted	0.076	0.045	0.044	0.036	0.047	20.973	**p<0.001**	n=9898
Adjusted for socio-demographic covariates	0.075	0.046	0.043	0.037	0.047	11.013	**p= 0.026**	n=9898
Adjusted for socio-demographic covariates and baseline disease	0.076	0.056	0.042	0.037	0.054	7.779	p= 0.100	n=7320
Morbid Obesity (BMI ≥ 40)								
Unadjusted	0.073	0.019	0.029	0.017	0.041	54.359	**p<0.001**	n=9898
Adjusted for socio-demographic covariates	0.075 [a,b,c,d]	0.018 [a]	0.028 [b]	0.016 [c,e]	0.040 [d,e]	78.263	**p<0.001**	n=9898

Note: Socio-demographic covariates adjusted for include: sex, age, own occupation, education level and monthly household income quintile.
[a,b,c,d,e] indicate statistically significant pairwise differences in disease prevalence at Bonferroni corrected significance level of p=0.005.
Bold values indicate a statistically significant omnibus Wald's test at significance level of p=0.05.

References

Appleton, K., Krumplevska, K., Smith, E., Rooney, C., McKinley, M. & Woodside, J. 2018. Low fruit and vegetable consumption is associated with low knowledge of the details of the 5-a-day fruit and vegetable message in the UK: findings from two cross-sectional questionnaire studies. *Journal of human nutrition and dietetics,* 31**,** 121-130.

Asparouhov, T. & Muthén, B. 2014. Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus web notes,* 21**,** 1-22.

Banks, E., Joshy, G., Korda, R. J., Stavreski, B., Soga, K., Egger, S., Day, C., Clarke, N. E., Lewington, S. & Lopez, A. D. 2019. Tobacco smoking and risk of 36 cardiovascular disease subtypes: fatal and non-fatal outcomes in a large prospective Australian study. *BMC medicine,* 17**,** 1-18.

Barta, J. A., Powell, C. A. & Wisnivesky, J. P. 2019. Global epidemiology of lung cancer. *Annals of global health,* 85.

Bécares, L. 2021. Health and socio-economic inequalities by sexual orientation among older women in the United Kingdom: findings from the UK household longitudinal study. *Ageing & Society,* 41**,** 2416-2434.

Ben-Shlomo, Y. & Kuh, D. 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. Oxford University Press.

Benzies, K. M., Wångby, M. & Bergman, L. R. 2008. Stability and change in health-related behaviors of midlife Swedish women. *Health Care for Women International,* 29**,** 997-1018.

Birch, J., Petty, R., Hooper, L., Bauld, L., Rosenberg, G. & Vohra, J. 2019. Clustering of behavioural risk factors for health in UK adults in 2016: a cross-sectional survey. *Journal of Public Health,* 41**,** e226-e236.

Bolin, K. & Lindgren, B. 2016. Non-monotonic health behaviours–implications for individual health-related behaviour in a demand-for-health framework. *Journal of Health Economics,* 50**,** 9-26.

Botoseneanu, A. & Liang, J. 2012. The effect of stability and change in health behaviors on trajectories of body mass index in older Americans: a 14-year longitudinal study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences,* 67**,** 1075-1084.

Britton, A., Ben-Shlomo, Y., Benzeval, M., Kuh, D. & Bell, S. 2015. Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies. *BMC medicine,* 13**,** 1-9.

Burns, D. M. 2003. Epidemiology of smoking-induced cardiovascular disease. *Progress in cardiovascular diseases,* 46**,** 11-29.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E. & Pan, Y. 2014. Early childhood investments substantially boost adult health. *Science,* 343**,** 1478-1485.

Chaoqun, D., Hua, C., Yeqin, Y., Yi, L., Yumei, S. & Hongyu, S. 2022. Patterns of risky health behaviors and associations with chronic diseases among young adult nursing students: a latent class analysis. *Journal of Nursing Research,* 30**,** e243.

Chen, Q., Luo, W., Palardy, G. J., Glaman, R. & McEnturff, A. 2017. The efficacy of common fit indices for enumerating classes in growth mixture models when nested data structure is ignored: A Monte Carlo study. *Sage Open,* 7**,** 2158244017700459.

Chuang, Y.-C., Chuang, K.-Y. & Yang, T.-H. 2013. Social cohesion matters in health. *International journal for equity in health,* 12**,** 1-12.

Cleland, C., Ferguson, S., Ellis, G. & Hunter, R. F. 2018. Validity of the International Physical Activity Questionnaire (IPAQ) for assessing moderate-to-vigorous physical activity and sedentary behaviour of older adults in the United Kingdom. *BMC medical research methodology,* 18**,** 1-12.

Conry, M. C., Morgan, K., Curry, P., McGee, H., Harrington, J., Ward, M. & Shelley, E. 2011. The clustering of health behaviours in Ireland and their relationship with mental health, self-rated health and quality of life. *BMC public health,* 11**,** 1-10.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A. & Sallis, J. F. 2003. International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise,* 35**,** 1381-1395.

Daw, J., Margolis, R. & Wright, L. 2017. Emerging adulthood, emergent health lifestyles: Sociodemographic determinants of trajectories of smoking, binge drinking, obesity, and sedentary behavior. *Journal of health and social behavior,* 58**,** 181-197.

de Almeida, M. G. N., Nascimento-Souza, M. A., Lima-Costa, M. F. & Peixoto, S. V. 2020. Lifestyle factors and multimorbidity among older adults (ELSI-Brazil). *European Journal of Ageing,* 17**,** 521-529.

Divo, M. J., Martinez, C. H. & Mannino, D. M. 2014. Ageing and the epidemiology of multimorbidity. *European Respiratory Journal,* 44**,** 1055-1068.

Dumith, S. C., Muniz, L. C., Tassitano, R. M., Hallal, P. C. & Menezes, A. M. 2012. Clustering of risk factors for chronic diseases among adolescents from Southern Brazil. *Preventive medicine,* 54**,** 393-396.

Fortin, M., Haggerty, J., Almirall, J., Bouhali, T., Sasseville, M. & Lemieux, M. 2014. Lifestyle factors and multimorbidity: a cross sectional study. *BMC public health,* 14**,** 1-8.

Foster, H. M., Celis-Morales, C. A., Nicholl, B. I., Petermann-Rocha, F., Pell, J. P., Gill, J. M., O'Donnell, C. A. & Mair, F. S. 2018. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. *The Lancet Public Health,* 3**,** e576-e585.

Freisling, H., Viallon, V., Lennon, H., Bagnardi, V., Ricci, C., Butterworth, A. S., Sweeting, M., Muller, D., Romieu, I. & Bazelle, P. 2020. Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. *BMC medicine,* 18**,** 1-11.

French, S., Rosenberg, M. & Knuiman, M. 2008. The clustering of health risk behaviours in a Western Australian adult population. *Health promotion journal of Australia,* 19**,** 203-209.

Fromme, K., Corbin, W. R. & Kruse, M. I. 2008. Behavioral risks during the transition from high school to college. *Developmental psychology,* 44**,** 1497.

Fumagalli, L., Knies, G. & Buck, N. 2017. Understanding Society, The UK Household Longitudinal Study, Harmonised British Household Panel Survey (BHPS) User Guide.

Graham, H., Hutchinson, J., Law, C., Platt, L. & Wardle, H. 2016. Multiple health behaviours among mothers and partners in England: clustering, social patterning and intra-couple concordance. *SSM-population health,* 2**,** 824-833.

Hailey, V., Bloomberg, M., Hamer, M. & Fisher, A. 2023. Association between neighbourhood cohesion and physical activity trajectories during the COVID-19 pandemic using data from Understanding Society: The UK Household Longitudinal Study & COVID-19 sub-study. *Preventive Medicine Reports,* 35**,** 102392.

He, L., Biddle, S. J., Lee, J. T., Duolikun, N., Zhang, L., Wang, Z. & Zhao, Y. 2021. The prevalence of multimorbidity and its association with physical activity and sleep duration in middle aged and elderly adults: a longitudinal analysis from China. *International Journal of Behavioral Nutrition and Physical Activity,* 18**,** 77.

Head, A., Fleming, K., Kypridemos, C., Schofield, P. & O'Flaherty, M. 2020. Dynamics of multimorbidity in England between 2004 and 2019: a descriptive epidemiology study. *European Journal of Public Health,* 30**,** ckaa165. 818.

Hiscock, R., Bauld, L., Amos, A., Fidler, J. A. & Munafò, M. 2012. Socioeconomic status and smoking: a review. *Annals of the New York Academy of Sciences,* 1248**,** 107-123.

Ho, I. S.-S., Azcoaga-Lorenzo, A., Akbari, A., Davies, J., Hodgins, P., Khunti, K., Kadam, U., Lyons, R., McCowan, C. & Mercer, S. W. 2022. Variation in the estimated prevalence of multimorbidity: systematic review and meta-analysis of 193 international studies. *BMJ open,* 12**,** e057017.

Hurcombe, R., Bayley, M. & Goodman, A. 2010. Ethnicity and alcohol: A review of the UK literature.

Ingram, E., Ledden, S., Beardon, S., Gomes, M., Hogarth, S., McDonald, H., Osborn, D. P. & Sheringham, J. 2021. Household and area-level social determinants of multimorbidity: a systematic review. *J Epidemiol Community Health,* 75**,** 232-241.

Institute for Social and Economic Research & Essex, U. o. 2023. Understanding Society: Waves 1-13, 2009-2022 and harmonised BHPS: Waves 1-18, 1991-2009. UK Data Service.

IPAQ Research Committee 2005. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms. *http://www. ipaq. ki. se/scoring. pdf*.

Jayes, L., Haslam, P. L., Gratziou, C. G., Powell, P., Britton, J., Vardavas, C., Jimenez-Ruiz, C., Leonardi-Bee, J., Dautzenberg, B. & Lundbäck, B. 2016. SmokeHaz: systematic reviews and meta-analyses of the effects of smoking on respiratory health. *Chest,* 150**,** 164-179.

Karamouzian, M., Cui, Z., Hayashi, K., Debeck, K., Reddon, H., Buxton, J. A. & Kerr, T. 2024. Longitudinal polysubstance use patterns and non-fatal overdose: a repeated measures latent class analysis. *International Journal of Drug Policy***,** 104301.

Katikireddi, S. V., Skivington, K., Leyland, A. H., Hunt, K. & Mercer, S. W. 2017. The contribution of risk factors to socioeconomic inequalities in multimorbidity across the lifecourse: a longitudinal analysis of the Twenty-07 cohort. *BMC medicine,* 15**,** 1-10.

Kingston, A., Comas-Herrera, A. & Jagger, C. 2018. Forecasting the care needs of the older population in England over the next 20 years: estimates from the Population Ageing and Care Simulation (PACSim) modelling study. *The Lancet Public Health,* 3**,** e447-e455.

Knies, G. & Kumari, M. 2022. Multimorbidity is associated with the income, education, employment and health domains of area-level deprivation in adult residents in the UK. *Scientific Reports,* 12**,** 7280.

Lanza, S. T. & Cooper, B. R. 2016. Latent class analysis for developmental research. *Child Development Perspectives,* 10**,** 59-64.

Lanza, S. T., Tan, X. & Bray, B. C. 2013. Latent class analysis with distal outcomes: A flexible model-based approach. *Structural equation modeling: a multidisciplinary journal,* 20**,** 1-26.

Lawder, R., Harding, O., Stockton, D., Fischbacher, C., Brewster, D. H., Chalmers, J., Finlayson, A. & Conway, D. I. 2010. Is the Scottish population living dangerously? Prevalence of multiple risk factors: the Scottish Health Survey 2003. *BMC Public Health,* 10**,** 1-13.

Lear, S. A., Hu, W., Rangarajan, S., Gasevic, D., Leong, D., Iqbal, R., Casanova, A., Swaminathan, S., Anjana, R. M. & Kumar, R. 2017. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. *The Lancet,* 390**,** 2643-2654.

Leventhal, A. M., Huh, J. & Dunton, G. F. 2014. Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis. *Perspectives in Public Health,* 134**,** 331-338.

Mahalik, J. R., Levine Coley, R., McPherran Lombardi, C., Doyle Lynch, A., Markowitz, A. J. & Jaffee, S. R. 2013. Changes in health risk behaviors for males and females from early adolescence through early adulthood. *Health psychology,* 32**,** 685.

Mair, F. S. & Jani, B. D. 2020. Emerging trends and future research on the role of socioeconomic status in chronic illness and multimorbidity. *The Lancet Public Health,* 5**,** e128-e129.

Marmot, M., Allen, J., Bell, R., Bloomer, E. & Goldblatt, P. 2012. WHO European review of social determinants of health and the health divide. *The Lancet,* 380**,** 1011-1029.

Mawditt, C., Sacker, A., Britton, A., Kelly, Y. & Cable, N. 2016. The clustering of health-related behaviours in a British population sample: testing for cohort differences. *Preventive medicine,* 88**,** 95-107.

McFall, S., Nandi, A. & Platt, L. 2017. Understanding society: UK household longitudinal study: user guide to ethnicity and immigration research. *Institute for Social and Economic Research*.

McGeoch, L. J., Ross, S., Massa, M. S., Lewington, S. & Clarke, R. 2023. Cigarette smoking and risk of severe infectious respiratory diseases in UK adults: 12-year follow-up of UK biobank. *Journal of Public Health,* 45**,** e621-e629.

Meader, N., King, K., Moe-Byrne, T., Wright, K., Graham, H., Petticrew, M., Power, C., White, M. & Sowden, A. J. 2016. A systematic review on the clustering and co-occurrence of multiple risk behaviours. *BMC public health,* 16**,** 1-9.

Meier, P. S., Holmes, J., Brennan, A. & Angus, C. 2021. Alcohol policy and gender: a modelling study estimating gender-specific effects of alcohol pricing policies. *Addiction,* 116**,** 2372-2384.

Meier, P. S., Meng, Y., Holmes, J., Baumberg, B., Purshouse, R., Hill-McManus, D. & Brennan, A. 2013. Adjusting for unrecorded consumption in survey and per capita sales data: quantification of impact on gender-and age-specific alcohol-attributable fractions for oral and pharyngeal cancers in Great Britain. *Alcohol and Alcoholism,* 48**,** 241-249.

Mercer, S., Furler, J., Moffat, K., Fischbacher-Smith, D. & Sanci, L. 2016. *Multimorbidity: technical series on safer primary care*, World Health Organization.

Mishra, V. K., Srivastava, S. & Murthy, P. 2021. Population attributable risk for multimorbidity among adult women in India: Do smoking tobacco, chewing tobacco and consuming alcohol make a difference? *PloS one,* 16**,** e0259578.

Muthén, B. & Muthén, L. 2017. Handbook of item response theory. Chapman and Hall/CRC Boca Raton, FL, USA:.

National Institute for Health Clinical Excellence 2022. *Alcohol-use disorders: diagnosis and clinical management of alcohol-related physical complications*, National Institute for Health and Clinical Excellence (NICE).

Noble, N., Paul, C., Turon, H. & Oldmeadow, C. 2015. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. *Preventive medicine,* 81**,** 16-41.

Nylund-Gibson, K. & Choi, A. Y. 2018. Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science,* 4**,** 440.

Ocean, N., Howley, P. & Ensor, J. 2019. Lettuce be happy: A longitudinal UK study on the relationship between fruit and vegetable consumption and well-being. *Social Science & Medicine,* 222**,** 335-345.
Opazo Breton, M., Gillespie, D., Pryce, R., Bogdanovica, I., Angus, C., Hernandez Alava, M., Brennan, A. & Britton, J. 2022. Understanding long-term trends in smoking in England, 1972–2019: an age–period–cohort approach. *Addiction,* 117**,** 1392-1403.
Pandey, R., Kumar, N., Paroha, S., Prasad, R., Yadav, M., Jain, S. & Yadav, H. 2013. Impact of obesity and diabetes on arthritis: An update. *Health,* 5**,** 143.
Pannu, G., Zaman, S., Bhala, N. & Zaman, R. 2009. Alcohol use in South Asians in the UK. British Medical Journal Publishing Group.
Pastuszak, A., Lisowski, K., Lewandowska, J. & Buśko, K. 2014. Level of physical activity of physical education students according to criteria of the IPAQ questionnaire and the recommendation of WHO experts. *Biomedical Human Kinetics,* 6.
Porcu, M. & Giambona, F. 2017. Introduction to latent class analysis with applications. *The Journal of Early Adolescence,* 37**,** 129-158.
Rauber, F., Louzada, M. L. d. C., Steele, E. M., Millett, C., Monteiro, C. A. & Levy, R. B. 2018. Ultra-processed food consumption and chronic non-communicable diseases-related dietary nutrient profile in the UK (2008–2014). *Nutrients,* 10**,** 587.
Schuit, A. J., van Loon, A. J. M., Tijhuis, M. & Ocké, M. C. 2002. Clustering of lifestyle risk factors in a general adult population. *Preventive medicine,* 35**,** 219-224.
Shan, J., Yin, R. & Panuthai, S. 2024. Body mass index and multimorbidity risk: A systematic review and dose-response meta-analysis. *Archives of Gerontology and Geriatrics***,** 105418.
Shang, X., Peng, W., Wu, J., He, M. & Zhang, L. 2020. Leading determinants for multimorbidity in middle-aged Australian men and women: A nine-year follow-up cohort study. *Preventive Medicine,* 141**,** 106260.
Sinha, P., Calfee, C. S. & Delucchi, K. L. 2021. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Critical care medicine,* 49**,** e63-e79.
Skou, S. T., Mair, F. S., Fortin, M., Guthrie, B., Nunes, B. P., Miranda, J. J., Boyd, C. M., Pati, S., Mtenga, S. & Smith, S. M. 2022. Multimorbidity. *Nature Reviews Disease Primers,* 8**,** 48.
Suhag, A., Webb, T. L. & Holmes, J. 2024. Longitudinal clustering of health behaviours and their association with multimorbidity in older adults in England: A latent class analysis. *Plos one,* 19**,** e0297422.
Tegegne, T. K., Islam, S. M. S. & Maddison, R. 2022. Effects of lifestyle risk behaviour clustering on cardiovascular disease among UK adults: latent class analysis with distal outcomes. *Scientific Reports,* 12**,** 17349.
van den Broek, T. 2021. Early-life circumstances, health behavior profiles, and later-life health in Great Britain. *Journal of Aging and Health,* 33**,** 317-330.
Van Gaal, L. F., Mertens, I. L. & De Block, C. E. 2006. Mechanisms linking obesity with cardiovascular disease. *Nature,* 444**,** 875-880.
Wadd, S. & Papadopoulos, C. 2014. Drinking behaviour and alcohol-related harm amongst older adults: analysis of existing UK datasets. *BMC research notes,* 7**,** 1-9.
Wagner, C., Carmeli, C., Jackisch, J., Kivimäki, M., van der Linden, B. W., Cullati, S. & Chiolero, A. 2024. Life course epidemiology and public health. *The Lancet Public Health,* 9**,** e261-e269.
Whitaker, V., Oldham, M., Angus, C., Fairbrother, H. & Holmes, J. 2020. The problem with good news: how should public health actors respond when alcohol consumption declines? *Addiction (Abingdon, England),* 115**,** 1987.

Wirt, A. & Collins, C. E. 2009. Diet quality–what is it and does it matter? *Public health nutrition,* 12**,** 2473-2492.

Wister, A. V., Levasseur, M., Griffith, L. E. & Fyffe, I. 2015. Estimating multiple morbidity disease burden among older persons: a convergent construct validity study to discriminate among six chronic illness measures, CCHS 2008/09. *BMC geriatrics,* 15**,** 1-12.

Extended Statistical Analysis on the UKHLS dataset

Given the significantly higher prevalence of most health conditions in Wave 7 compared to Wave 13, likely due to methodological differences in data collection between the two waves, we adopted an alternative approach to examine health outcomes.

Instead of analysing whether participants had ever reported a condition, we focused on whether they reported still having it. For Wave 7, this involved tracking participants from the wave they joined the UKHLS to determine if they still had the condition they reported at the time they joined. This approach was necessary because only new entrants (i.e., those not previously interviewed in UKHLS) were asked to report all conditions they had ever been diagnosed with and whether they still had them, with exceptions for stroke and heart attack, where no follow-up questions were asked. As these participants transitioned into old entrants (those interviewed in earlier waves), we also monitored whether they reported any new conditions in subsequent waves and if those persisted. Since old entrants were not asked follow-up questions about conditions reported in previous waves, we had no way to assess changes in those conditions. Therefore, we calculated disease presence by tracing back from Wave 7 to Wave 1 for the presence or absence of disease.

For Wave 13, we assessed disease prevalence by tracing back from Wave 10, when all participants were asked anew whether they had ever had a condition and if they still had it. From Wave 11 onwards, all participants, regardless of when they joined, were asked to report whether they still had previously disclosed conditions or had developed new ones. However, conditions such as stroke, heart attack, multiple sclerosis, and any conditions categorised as "other" were not followed up from previous waves. Thus, the presence of these conditions in Wave 13 was recorded only if they were reported as new in that wave. For all other conditions,

presence was recorded only if participants claimed to still have the condition (whether newly diagnosed or carried forward from a previous wave) in Wave 13.

The descriptive statistics of prevalence calculated in this manner are presented in Table A1.

Table A1. Prevalence of chronic conditions at baseline Wave 7 and Wave 13

		Sample (n=18008)
Disease status at Wave 7		N (%)
Multimorbidity (>= 2 chronic conditions)	Has condition	2534 (14.1)
	Missing	3320 (18.4)
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)	Has condition	2802 (15.6)
	Missing	4302 (23.9)
Cancer (All types)	Has condition	298 (1.7)
	Missing	4617 (25.6)
Obesity-related (Diabetes, High Blood pressure)	Has condition	3291 (18.3)
	Missing	4073 (22.6)
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)	Has condition	1778 (9.9)
	Missing	4553 (25.3)
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)	Has condition	902 (5.0)
	Missing	4596 (25.5)
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)	Has condition	957 (5.3)
	Missing	4588 (25.5)
Disease status at Wave 13		
Multimorbidity (>= 2 chronic conditions)	Has condition	1408 (7.8)
	Missing	170 (0.9)
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)	Has condition	1344 (7.5)
	Missing	3587 (19.9)
Cancer (All types)	Has condition	121 (0.7)
	Missing	4766 (26.5)
Obesity-related (Diabetes, High Blood pressure)	Has condition	2086 (11.6)
	Missing	2726 (15.1)
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)	Has condition	935 (5.2)
	Missing	3999 (22.2)
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)	Has condition	469 (2.6)
	Missing	170 (0.9)
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)	Has condition	1017 (5.6)
	Missing	170 (0.9)
Morbid Obesity (BMI ≥40)	Has condition	640(3.4)
	Missing	103(0.4)

Health behaviour clusters and disease status

After controlling for both sociodemographic characteristics and respective diseases at baseline, the prevalence of multimorbidity, respiratory disorders, arthritis, cardiovascular diseases, obesity-related conditions, cancers, morbid obesity and conditions classified as 'other' differed significantly across clusters — omnibus Wald test χ^2(df = 6) multimorbidity = 37.61, p <0.001; χ^2(df = 6) respiratory disorders = 37.91, p<0.001; χ^2(df = 6) Arthritis = 28.00, p<0.001; and χ^2(df = 6) cardiovascular diseases = 14.40, p =0.026; χ^2(df = 6) obesity-related conditions =15.02, p=0.02, χ^2(df = 6) cancers =12354, p<0.001, χ^2(df = 6) morbid obesity =84.12, p<0.001, χ^2(df = 6)other conditions =16.10, p=0.013, respectively. To further investigate which clusters differed in the proportions of participants with each disease profile, we conducted pairwise comparisons between clusters, with results adjusted for disease status and socio-demographic variables as shown in Table A2.

Similar to findings in the original analysis where health conditions were assessed based on whether participants had 'ever' experienced them, rather than their current status as in the present analysis — our data revealed an unexpected trend: Participants characterised as *Overall low-risk* had a significantly higher prevalence of multimorbidity, obesity-related conditions, and arthritis compared to participants in the *Hazardous drinkers, Insufficiently active with Poor diet* and *Insufficiently Active* clusters. Participants characterised as *Overall low-risk* also had a higher prevalence of morbid obesity compared to all other clusters except the cluster labelled *Smokers and Drinkers*. Notably, the higher prevalence of multimorbidity in the *Overall low-risk* persisted after adjusting for sociodemographic factors. This unexpected finding is consistent with observations from our original analysis.

Participants characterised as *Smokers* had a higher prevalence of respiratory disorders than most other clusters. *Smoker and Drinkers* had the second highest prevalence of respiratory disorders; however, significant pairwise differences were not detected when compared to other clusters after applying the Bonferroni correction. Unexpectedly, participants classified as *Overall low-risk* had a higher prevalence of respiratory disorders than those in the *Insufficiently Active, Insufficiently Active, Poor Diet,* and *Hazardous drinkers, Insufficiently active with Poor diet* clusters. Notably, the proportional distribution of respiratory conditions align with the original analysis; however, in the current analysis, the pairwise differences reached significance, likely due to the overall higher prevalence observed.

Smokers and Drinkers exhibited a significantly lower prevalence of cancer than all other clusters. This lower prevalence of cancer may be partially attributed to the fact that this cluster had the highest proportion of individuals in the youngest age groups (between 16 and 34 years). Yet, the proportional distribution of cancer prevalence across clusters is consistent with the findings from the original analysis.

Finally, differences in the prevalence of cardiovascular diseases and conditions classified as 'other' were observed across clusters, though these differences did not reach significance at the Bonferroni-corrected level in pairwise comparisons. Nevertheless, we did observe that participants characterised as *Smokers* and *Overall low-risk* had a higher prevalence of both cardiovascular diseases and other diseases compared to other clusters.

Overall, these findings parallel our original analysis in the proportional distribution of prevalence across clusters. However, a greater number of pairwise differences reached

significance at the Bonferroni-corrected level in this analysis, likely due to the higher overall prevalence observed.

Table A2. Pairwise comparisons of disease status prevalence across clusters

Disease outcomes (Wave 13)	Overall low-risk	Insufficiently Active	Poor Diet, Insufficiently active	Hazardous and harmful drinkers	Hazardous drinkers, Insufficiently active with Poor diet	Smokers and Drinkers	Smokers	Wald's omnibus test χ^2(df = 6)	Wald's omnibus p-value	Sample size
	19.97%	**18.29%**	**23.22%**	**10.72%**	**14.03%**	**5.09%**	**8.68%**			
Multimorbidity (>= 2 chronic conditions)										
Unadjusted	0.156 g,h,i,j,k	0.047 f,j	0.064 a,d,i	0.074 e,k	0.033 a,c,e	0.043 b,g,h	0.109 b,c,d,f	164.00	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.146 e,f,g,h	0.052 d,h	0.063 b,g	0.077 c	0.036 a,c,f	0.046 e	0.107 a,b,d	81.90	**<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.151 b,c,d	0.065 d	0.07 b	0.088	0.043 a,c	0.055	0.122 a	37.61	**<0.001**	n=14688
Cancers (All types)										
Unadjusted	0.009 a	0.009 b	0.009 c	0.013 d	0.005 e	0.000 a,b,c,d,e,f	0.017 f	111.95	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.009 a	0.011 b	0.010 c	0.015 d	0.006 e	0.000 a,b,c,d,e,f	0.020 f	16760	**<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.008 a	0.013 b	0.011 c	0.017 d	0.006 e	0.000 a,b,c,d,e,f	0.025 f	12354	**<0.001**	n=13391
Obesity-related (Diabetes, High Blood pressure)										
Unadjusted	0.217 a,b,c,d,e,f	0.095 a,g	0.114 b	0.132 c	0.087 d,h	0.091 e	0.155 f,g,h	111.57	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.213 a,b,c,d,	0.104 a	0.115 b	0.136	0.093 c	0.096 d	0.156	50.52	**<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.215 a	0.124	0.129	0.153	0.097 a	0.102	0.167	15.02	**0.020**	n=13935
Respiratory disorders (Asthma, Emphysema, Chronic Bronchitis)										
Unadjusted	0.112 d,e,f	0.036 c,f	0.05 b,e	0.071	0.044 a,d	0.113	0.135 a,b,c	87.54	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.112 g,h,i	0.040 e,f,i	0.051 b,d,h	0.076	0.046 a,c,g	0.118 a,b,e	0.139 c,d,f	62.95	**<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.113 d,e,f	0.050 a,f	0.053 b,e	0.080	0.048 c,d	0.128	0.142 a,b,c	37.91	**<0.001**	n=13455
Arthritis (Osteoarthritis, Rheumatoid Arthritis, Other)										
Unadjusted	0.173 a,b,c,d,e	0.048 a,f,g	0.087 b,g,j	0.084 c	0.052 d,h,j	0.044 e,i	0.127 f,h,i	129.53	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.166 a,b,c	0.061 a,d,e	0.088 b,f	0.095 d	0.059 c,g	0.051	0.130 e,f,g	48.26	**<0.001**	n=18008

*Adjusted for socio-demographic covariates and baseline disease	0.177 [a,b]	0.073 [a,c]	0.102	0.116 [c]	0.064 [b]	0.060	0.140	28.00	**0.0001**	n=13706
Cardiovascular diseases (Congestive Heart Failure, Coronary Heart Disease, Angina, Heart Attack, Stroke)										
Unadjusted	0.061 [a,b,c,d]	0.017 [a]	0.020 [b]	0.032 [c,e]	0.009 [d,e,f]	0.017	0.04 [f]	64.68	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.057 [a,b,c]	0.019 [a]	0.020 [b]	0.032	0.010 [c,d]	0.018	0.039 [d]	35.54	**<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.060	0.022	0.021	0.034	0.014	0.025	0.040	14.40	**0.026**	n=13412
Other (Any Liver Condition, Hypothyroidism, Multiple Sclerosis, Any other long-standing conditions)										
Unadjusted	0.088**[a,b,c,d,e]**	0.051 **[a]**	0.055 **[b,g]**	0.027 **[c,f,g]**	0.037 **[d,h]**	0.024 **[e]**	0.074 **[f,h]**	59.74	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.084 **[a,b,c]**	0.054	0.053 **[a]**	0.029 **[b,d]**	0.038 **[c]**	0.025	0.073 **[d]**	30.22	**<0.001**	n=18008
*Adjusted for socio-demographic covariates and baseline disease	0.085	0.058	0.055	0.035	0.040	0.028	0.081	16.10	**0.013**	n=13420
Morbid Obesity (BMI ≥40)										
Unadjusted	0.076 **[a,b,c,d]**	0.018 **[a,e]**	0.031 **[b]**	0.029 **[c]**	0.018 **[d,f]**	0.036	0.051 **[e,f]**	78.26	**<0.001**	n=18008
Adjusted for socio-demographic covariates	0.077**[a,b,c,d,e]**	0.018 **[a]**	0.031 **[b]**	0.029 **[c]**	0.018 **[d]**	0.036	0.050 **[e]**	84.12	**<0.001**	n=18008

Note: Socio-demographic covariates adjusted for include sex, age, own occupation, education level and monthly household income quintile.
[a,b,c,d,e] indicates statistically significant pairwise differences in disease prevalence at Bonferroni corrected significance level of p=0.002.
Bold values indicate a statistically significant omnibus Wald's test at significance level of p=0.05

Chapter 6: Moderating effects of socio-demographic characteristics on the relationship between risk behaviour clusters and multimorbidity in UK older adults

Overview: This paper presents a study using latent class moderation to examine whether socio-demographic characteristics influence the relationship between risk behaviour clusters and multimorbidity in older adults. Utilising data from the ELSA study, earlier research found that health-risk behaviour profiles did not fully explain the variance in health outcomes across different clusters. This prompted further investigation into whether socio-demographic factors could moderate the relationship between these behaviour clusters and disease status. Unlike previous studies that controlled for these characteristics, latent class moderation allows for exploring the combined effects of socio-demographic factors and multiple risk behaviour patterns over time (i.e. latent class membership) on disease. The methodology adopted in this study offers a novel way to examine whether multiple risk behaviours and socio-demographic characteristics interact to impact health outcomes in older adults.

Introduction

Multimorbidity, or the simultaneous presence of two or more chronic conditions, accounts for a significant proportion of healthcare spending in the UK and is strongly associated with physical decline, mortality, and reduced quality of life (Soley-Bori et al., 2021). Increasing age is a well-established risk factor for multimorbidity (Willadsen et al., 2016). By 2035, over two-thirds of the population aged 65 and over in England could be affected by multimorbidity (Head

et al., 2021b). However, in recent cohorts, multimorbidity is not only becoming more prevalent but is also manifesting at earlier ages — a trend that is likely to continue over the next two decades in England (Head et al., 2020; Kingston, Comas-Herrera and Jagger, 2018). In addition to age, multimorbidity is significantly influenced by socio-economic deprivation. For example, a UK study found that the onset of multimorbidity occurred 10-15 years earlier in people living in the most deprived areas compared with the most affluent (Barnett et al., 2012). Furthermore, socio-economic disparities in multimorbidity prevalence in England have widened from 2004 to 2019, with the most notable increases observed among working-age adults (Head et al., 2021b). This persistent rise in multimorbidity rates and its uneven distribution across the population underscores the urgent need for preventive measures to delay its onset.

Lifestyle factors, including smoking, unhealthy nutrition, alcohol consumption, and physical inactivity—commonly known as SNAP behaviours—are risk factors common to several chronic diseases and multimorbidity (Willadsen et al., 2016). Targeting multiple modifiable risk behaviours holds promise not only for preventing or delaying the onset of multimorbidity but also for compressing severe illness and disability into a shorter period at later life stages (Bricca et al., 2023). However, a common challenge raised by intervention developers is that targeting multiple risk behaviours is difficult due to the 'infinite combinations of risk factors with which patients may present' (Prochaska et al., 2010). Thus, our previous work examined whether and how risk behaviours cluster over time and their link to disease outcomes (Suhag, Webb and Holmes, 2024). Our current aim is to investigate how socio-demographic characteristics moderate the relationship between the previously identified clusters of health risk behaviours and disease outcomes among older adults in England.

In epidemiological studies, behavioural risk factors are typically examined individually or aggregated into a cumulative risk index, which sums up the number of risky behaviours an

individual engages in. However, examining isolated risk behaviours overlooks the growing body of evidence suggesting that behavioural risk factors tend to co-occur in individuals and cluster in specific subgroups (Noble et al., 2015). This is crucial because combinations of risk behaviours can have synergistic health effects that can disproportionately increase a person's likelihood of experiencing negative health outcomes (Tegegne, Islam and Maddison, 2022). Similarly, the cumulative risk index has some key limitations: i) it does not tell us which combinations of behaviours drive the risk, ii) it treats multiple risk factors as interchangeable, suggesting that the simple fact of exposure to a risk matters more than the nature of that risk, and iii) it does not provide enough information about how lifestyles, i.e., combinations of multiple behavioural risk factors, vary across distinct subgroups. Combinations of risk behaviours are important because people cannot be accurately dichotomised into two distinct groups—those with healthy and unhealthy lifestyles. For instance, certain individuals who are physically active tend to consume alcohol more frequently than their inactive counterparts (e.g., young adults) (Leventhal, Huh and Dunton, 2014). Similarly, physical inactivity commonly co-occurred with a lack of fruit and vegetables in a study of older men in the UK (Zwolinsky, Raine and Robertson, 2016). Thus, health risk behaviours do not exhibit a monotonic relationship, meaning that as engagement in one behaviour increases, engagement in another behaviour may not consistently increase or decrease in a predictable manner. Rather, there are qualitative differences in behavioural profiles across different subgroups, which may present opportunities or problems when designing disease-prevention programs (Noble et al., 2015).

A promising approach for addressing the aforementioned gaps are clustering techniques such as Latent Class Analysis (LCA). LCA categorises individuals into homogeneous subgroups based on similarities in their characteristics (or response patterns to survey questions), thus helping to reduce data with patterns across many characteristics or measures into a more

parsimonious form – a single categorical variable that describes a small set of subgroups. Although LCA has been used to study health risk behaviours (Hutchesson et al., 2021; Kino, Bernabé and Sabbah, 2017; Irizar et al., 2022), its application in the context of chronic disease outcomes, especially using longitudinal data on risk behaviours, remains limited (Suhag, Webb and Holmes, 2024). Using latent class analysis to examine an individual's entire range of behavioural risk factors over several years allows us to observe and assess the health impact of cumulative lifestyle patterns (i.e. clusters) as well any patterns of change (i.e. abstaining from a risky behaviour or adopting healthier habits). It also allows us to examine participant characteristics associated with identified clusters and to identify how clusters are associated with disease outcomes. Equally, latent class analysis offers a promising method for studying the aetiology of chronic diseases characterised by complex causality, especially the role of social determinants of health (Hoekstra, 2013). For instance, if two risk factors cluster together with each other more than the other factors, this may suggest that they could be influenced by a common source. Similarly, if certain characteristics are more prevalent in certain subgroups (e.g., lower income), factors common to that subgroup (e.g., reduced access to affordable, healthy food) may give rise to the clustering of those risk factors and associated health outcomes.

Thus, in a previous study (Suhag, Webb and Holmes, 2024), we used repeated measures latent class analysis to examine how the SNAP behaviours cluster together over time (i.e. across an eight-year period) in certain subgroups. We also investigated associations between the identified subgroups, socio-demographic characteristics and disease outcomes. Specifically, we identified seven clusters of older adults with distinct patterns of behaviour and socio-demographic characteristics. The prevalence of multimorbidity, complex multimorbidity, and disorders of six body systems differed across the clusters. Specifically, the findings suggested that clusters characterised by health-promoting behaviours (i.e. adequate physical activity and

fruit and vegetable intake) had the lowest prevalence of all health conditions considered. However, the cluster characterised by physical inactivity (but no other risky behaviours) had a higher prevalence of multimorbidity, complex multimorbidity, and endocrine, nutritional, and metabolic disorders than the other clusters. In other words, we found that the relationship between risk behaviours and health outcomes is not a simple linear relationship between number of risk behaviours and adverse health outcomes (Myint et al., 2011).

The Present Research

The present work builds on our previous study (Suhag, Webb and Holmes, 2024) to examine whether and how socio-demographic characteristics moderate the relationship between patterns of SNAP behaviours over time and disease outcomes. In other words, rather than testing whether socio-demographic characteristics moderate the effect of single risk behaviours on disease outcomes (as traditional moderation analyses do; Luk and Tsoh, 2010; Cano et al., 2017; Kershaw et al., 2010), we use a multidimensional latent predictor, which represents subgroups of older adults with similar longitudinal patterns of SNAP behaviours. Using such a latent moderation model allows us to examine whether and how socio-demographic characteristics interact with a range of key modifiable behavioural risk factors for chronic disease—which not only co-occur in individuals but also tend to have synergistic health effects—to influence disease outcomes. In other words, such a moderation would allow us to examine how socio-demographic characteristics might amplify or attenuate the association between multiple risk behaviours over time and chronic diseases — an examination that is often neglected as socio-demographic characteristics are typically controlled for in studies investigating the influence of risk behaviours on disease (Artaud et al., 2016; Mishra, Srivastava and Murthy, 2021; He et al., 2021). Finally, using latent class moderation allows us to examine emergent properties arising from the interaction between socio-demographic

characteristics and clusters of health behaviour on disease status. Emergent properties refer to novel characteristics or patterns that emerge from the interaction of simpler components, often resulting in outcomes that cannot be predicted solely from the individual components (Desmond et al., 2019). Compared to the dominant epidemiological model, the latent class moderation approach is better suited to capture the possibility that the *determinants of disease rates* in populations may differ from the *determinants of which individuals develop disease* within a population (Diez Roux, 2004).

Thus, our aim is to use data from a longitudinal study of older adults to examine whether and how baseline socio-demographic characteristics moderate the relationship between previously identified clusters of SNAP risk behaviours over time (i.e. across an eight-year period) and disease status for multimorbidity, complex multimorbidity and six body system disorders.

Methods

Data

We used secondary data from the English Longitudinal Study of Ageing (ELSA), which is a nationally-representative panel study of community-dwelling adults aged 50 and over living in England (Banks, 2021). ELSA collects biennial data on mental and physical health, finances, and attitudes around ageing using computer-assisted interviews and questionnaires (Banks, 2021). For our study, we analysed data from 4759 respondents to the core questionnaire across six waves from Wave 4 (2008-2009) to Wave 9 (2018-2019). The sample selection and exclusion criteria are further described in our previous study (Suhag, Webb and Holmes, 2024).

Health Behaviour Measures

Data on health behaviours was taken from Waves 4-8. The process of choosing health-behaviour measures and arriving at the number of categories and cut-offs for those categories are described in Suhag, Webb and Holmes (2024). Briefly, we used a binary measure of current smoking status (smoker/non-smoker) for smoking behaviour. Portions of fruit and vegetable consumed per day were added together to create a single variable for each wave of data collection, which was subsequently divided into two categories (<5 or ≥5 portions per day). Alcohol consumption was categorised into four levels based on units (1 unit = 8g alcohol) consumed per week: harmful (>50 units for men, >35 units for women); hazardous (>14-50 units for men and >14-35 units for women); moderate (14 units or less); and abstainers (0 units) (Health, 2016). Following previous research (Dhalwani et al., 2016; Hamer, de Oliveira and Demakakos, 2014), we created a summary index to classify participants' levels of physical activity as: sedentary (no activity on a weekly basis); low (only mild activity at least once a week); moderate (moderate but no vigorous activity at least once a week); or high (any vigorous activity at least once a week).

Socio-demographic variables

Socio-demographic variables were taken from Wave 4 and included age, sex, occupation, education and wealth. For participants' own occupation, the three-class version of the National Statistics—Socio-economic Classification Scheme (McKnight and Elias, 2003) was used: professional and managerial occupations, intermediate occupations, and semi-routine and routine occupations. Education was taken from the IFS derived datafile provided with the ELSA dataset and grouped into 'degree/higher' (National Vocational Qualification NVQ4/NVQ5/degree or equivalent), 'intermediate' (higher education below degree, NVQ3/GCE A-level equivalent, NVQ2/GCE O-level equivalent, NVQ1/CSE other grade equivalent or foreign/other), 'no qualifications' (Fraser et al., 2014). Wealth was taken from the ELSA-derived dataset, which defines wealth as the net total non-pension wealth including property, possessions, housing, investments, savings, artwork, jewellery, and net of debt reported at the household level (i.e. an individual or a couple living at the same address who make joint financial decisions) (Tsimpida et al., 2022). Wealth was then grouped into tertiles to reduce measurement error and facilitate comparisons of health measures across equally sized groups within the population. All variables (except age) were converted into dummy variables, with the lowest category serving as the reference group.

Disease status

Information on 25 self-reported, physician-diagnosed physical and mental health conditions was collected at each wave, as listed in Table 1, and is used here to measure basic and complex multimorbidity (Singer et al., 2019a). We adopted the most widely cited and accepted definition of basic multimorbidity, which identifies it as involving two or more chronic conditions (Mercer et al., 2016). Consequently, we binary coded respondents as "yes" for

multimorbidity if they had two or more conditions from the 25-condition list and as "no" otherwise. Complex multimorbidity was defined as having three or more conditions affecting three or more body systems (Harrison et al., 2014). We based our selection of body system disorders (listed in Table 1) for calculating complex multimorbidity on a previous study using the ELSA dataset (Singer et al., 2019a). This study identified eight body systems subject to disorders as outlined by the International Classification of Diseases 10th Revision system: eye disorders; circulatory disorders; nervous disorders mental and behavioural problems; neoplasms; respiratory disorders; endocrine, nutritional and metabolic disorders; and musculoskeletal and connective system disorders. Using the self-reported presence or absence of three or more body system disorders, we derived a binary variable representing complex multimorbidity.

Table 1. Morbidities used to ascertain multimorbidity and complex multimorbidity

Body system disorders	Morbidities
1. Eye disorders	1. Glaucoma
	2. Macular degeneration
	3. Cataracts
2. Circulatory disorders	1. High blood pressure
	2. Angina
	3. Heart Attack
	4. Congestive heart failure
	5. Heart murmur
	6. Abnormal heart rhythm
	7. Stroke
3. Endocrine, nutritional and metabolic	1. Diabetic eye disease
	2. Diabetes
4. Musculoskeletal and connective system	1. Osteoporosis
	2. Arthritis
5. Respiratory	1. Lung disease
	2. Asthma
6. Neoplasms	1. Cancers
7. Nervous disorders	1. Parkinson's disease
	2. Alzheimer's disease
	3. Hallucinations
8. Mental and behavioural	1. Anxiety
	2. Depression
	3. Emotional problems
	4. Mood swings
	5 Dementia

Note. Adapted from Singer et al. 2019a

Statistical analysis

Data analysis was carried out in two phases. In the first phase, we explored how the SNAP behaviours cluster over time in older adults. This involved conducting a repeated measures latent class analysis (RMLCA) to estimate the best-fitting latent class model, where the latent classes represented subgroups of individuals with common patterns of SNAP behaviours over time. In the second phase, we investigated whether and how socio-demographic characteristics moderated the relationship between clusters of health behaviour over time and a range of disease outcomes. This phase involved incorporating auxiliary variables (i.e. moderators, covariates, and distal outcomes) into the latent class model to test for moderation.

In the first phase, we performed class enumeration of an unconditional RMLCA model (i.e. without any auxiliary variables). We identified the optimal number of classes representing heterogeneity in SNAP risk behaviours across five time points through iterative model fitting until a minimal increase in model fit or non-identification of the estimated model solution was found. Specifically, we created and compared latent class models with one to nine classes on a number of statistical criteria, classification diagnostics, and interpretability. These steps and its results are described in our previous study. But briefly, we found a seven-class model to best represent the data on SNAP behaviour patterns over time (Suhag, Webb and Holmes, 2024).

In the second phase, the seven-class latent model was extended to include auxiliary variables. Various methods exist for incorporating auxiliary variables into mixture models (e.g., one-step, classify-analyse, direct-inclusion), but we selected the maximum likelihood (ML) three-step manual approach, as it has been shown to reduce parameter shifts in the model and be less biased than other approaches (Zhu, Steele and Moustaki, 2017; Asparouhov and Muthén, 2014).

The maximum likelihood (ML) three-step manual approach involves the following steps: 1) perform class enumeration of the unconditional LCA model, 2) determine measurement error of the modal class assignment, and 3) specify the regression model(s) with the effect of auxiliary variables entered into the model. The first step involving class enumeration was described above in phase one of our study. In this step, we classified individuals into the latent class that they had the highest posterior probability of belonging to and assigned them to dummy-coded variables representing membership in one of the seven latent classes. In the second step, we extracted the logits for the classification probabilities of the modal class assignment (the class that an individual had the highest posterior probability of falling under) to be used in the final step to determine the measurement error associated with the modal class assignment. Finally, the dataset created in the first step (which includes modal class assignment) and the logits extracted in the second step were ready to be used in the third and final step: specifying the moderation model with auxiliary variables.

Separate moderation analyses were conducted for each socio-demographic variable (i.e. wealth, education, occupation, age, and sex) and each health outcome using logistic regression. For example, in the case of wealth, we used a logistic regression model to regress a given disease outcome (e.g. multimorbidity) on: i) wealth, ii) latent class membership, iii) the interaction between wealth and latent class membership, and iv) covariates. Covariates in this instance included the remaining socio-demographic variables (i.e. education, occupation, age, and sex) and disease status at baseline (multimorbidity disease status at the baseline Wave 4). The statistical significance of the interaction terms was used to address our main research question about moderation i.e. whether the relationship between clusters of health behaviour over time and disease outcomes was moderated by socio-demographic characteristics. To

account for the multiple tests, we adjusted the significance level using the Bonferroni correction ($\alpha = <0.005$). All analyses were done using the MPlus v8.5 software and R version v4.0.3 (Muthén and Muthén, 2017b; Nylund-Gibson and Choi, 2018).

Results

Parameter estimates for the moderation models are presented in Tables 2-6. Parameter estimates for the covariates included in each of these models are presented in the Supplementary Material (Table S1).

Wealth

There was a significant main effect of wealth on some disease outcomes (see Table X1). Belonging to the highest tertile of wealth was significantly associated with lower odds of being diagnosed with multimorbidity (OR=0.66, p<0.001), complex multimorbidity (OR=0.71, p<0.001), and endocrine disorders (OR=0.60, p<0.001), compared to the lowest tertiles of wealth. Belonging to the second tertile of wealth was also associated with lower odds of endocrine disorders (OR=0.59, p<0.001), compared to the lowest tertile. Latent class membership had a significant main effect on complex multimorbidity. Specifically, we found that compared to the *Low-risk* cluster, the *Abstainer but Inactive* cluster (OR=7.32, p<0.001) and *High-risk smokers* (OR=5.03, p<0.001) were seven to five times more likely, respectively, to have complex multimorbidity.

Education

In the moderation regression models with education as the primary predictor, we did not find a significant main effect for either education or latent classes on any disease outcome (see Table 3). Similarly, the moderation effects of education on latent class membership were not significant in these models.

Occupation

We did not find a significant main effect of occupation on disease outcomes (see Table 4). However, we found that belonging to the *High-risk smokers* cluster was associated with five

times higher odds of having respiratory disorders (OR=5.47, p<.001) compared to the *Low-risk* cluster. We did not find evidence that occupation moderated the relationship between latent classes and any of the disease outcomes.

Table 2. Tests of the moderating role of wealth in the relationship between risk behaviour latent class membership and disease outcomes

Disease outcomes		**Multimorbidity**		**Complex Multimorbidity**		**Circulatory disorders**		**Musculoskeletal disorders**		**Endocrine disorders**		**Respiratory disorders**		**Neoplasm**		**Eye disorders**	
		OR	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***
Intercept		0.02	<0.001	0.01	<0.001	0.10	<0.001	0.15	<0.001	0.27	0.08	0.01	<0.001	<0.001	<0.001	0.01	<0.001
Moderator (Wealth)																	
1st Tertile		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
2nd Tertile		0.82	0.05	0.93	0.47	0.87	0.15	0.98	0.86	**0.59***	**<0.001***	0.98	0.88	1.04	0.82	1.11	0.27
3rd Tertile		**0.66***	**<0.001***	**0.71***	**<0.001***	0.83	0.08	0.81	0.06	**0.60***	**<0.001***	1.07	0.68	0.73	0.13	0.99	0.88
Predictor (Latent class)																	
Low-risk		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive		1.83	0.14	3.10	0.02	1.29	0.56	1.18	0.67	2.07	0.18	4.34	0.06	0.96	0.96	1.86	0.09
Low-risk yet heavy		1.17	0.78	1.39	0.63	1.00	1.00	1.24	0.68	0.11	0.01	2.75	0.23	**<0.001***	**<0.001***	**3.89***	**<0.001***
Abstainers but inactive		1.59	0.22	**2.52***	**0.04***	0.97	0.93	0.87	0.67	2.71	0.04	4.26	0.05	1.14	0.84	1.96	0.05
Poor diet and inactive		1.53	0.29	1.79	0.23	1.41	0.42	1.05	0.89	1.18	0.79	2.65	0.19	1.69	0.42	1.89	0.09
Inactive, heavy drinkers		1.93	0.14	2.43	0.07	1.41	0.43	1.60	0.25	0.72	0.55	6.24	0.02	0.65	0.58	2.02	0.07
High-risk smokers		1.52	0.25	**2.65***	**0.03***	1.14	0.72	1.40	0.28	1.27	0.64	**10.38***	**<0.001***	0.68	0.58	1.62	0.15
Moderation (Latent class x Wealth)																	
Low-risk (Ref.)	1st Tertile	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive	2nd Tertile	1.11	0.84	0.35	0.05	1.04	0.94	1.10	0.84	1.22	0.81	0.42	0.34	0.95	0.96	0.64	0.32
	3rd Tertile	0.65	0.37	0.57	0.32	0.83	0.71	0.92	0.85	1.25	0.75	0.23	0.09	2.13	0.43	0.64	0.30
Low-risk yet heavy	2nd Tertile	1.25	0.72	0.35	0.18	1.16	0.80	1.02	0.98	13.59	0.02	0.38	0.36	****	****	0.22	0.01
	3rd Tertile	0.88	0.82	0.64	0.55	0.98	0.98	0.74	0.60	4.28	0.14	0.48	0.44	****	****	0.26	0.01
Abstainers but inactive	2nd Tertile	1.23	0.66	0.74	0.56	2.15	0.10	1.62	0.26	1.68	0.49	0.33	0.20	1.50	0.63	0.62	0.25
	3rd Tertile	1.11	0.83	0.75	0.61	1.41	0.46	0.77	0.55	1.11	0.89	0.61	0.57	1.10	0.92	0.75	0.49
Poor diet and inactive	2nd Tertile	0.96	0.93	0.61	0.38	0.81	0.67	0.93	0.87	1.88	0.47	0.32	0.25	0.94	0.94	0.73	0.49
	3rd Tertile	0.53	0.21	0.60	0.41	0.43	0.12	0.96	0.94	0.96	0.97	0.54	0.50	0.62	0.67	0.61	0.29
Inactive, heavy drinkers	2nd Tertile	1.13	0.81	0.51	0.24	0.95	0.93	0.68	0.44	2.45	0.28	0.12	0.03	2.95	0.25	0.54	0.18
	3rd Tertile	0.60	0.30	0.68	0.48	0.85	0.75	0.55	0.21	0.68	0.62	0.21	0.08	2.23	0.42	0.59	0.22
High-risk smokers	2nd Tertile	1.06	0.90	0.55	0.26	1.76	0.20	0.82	0.65	2.03	0.40	0.26	0.12	2.75	0.26	0.62	0.27
	3rd Tertile	0.60	0.30	0.68	0.48	0.85	0.75	0.55	0.21	0.68	0.62	0.21	0.08	2.23	0.42	0.59	0.22

Bold values are statistically significant at p ≤0.005

Table 3. Tests of the moderating role of education in the relationship between risk behaviour latent class membership and disease outcomes

Disease outcomes		**Multimorbidity**		**Complex Multimorbidity**		**Circulatory disorders**		**Musculoskeletal disorders**		**Endocrine disorders**		**Respiratory disorders**		**Neoplasm**		**Eye disorders**	
		OR	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***
Intercept		0.02	<0.001	0.01	<0.001	0.08	<0.001	0.02	<0.001	0.34	0.14	0.02	<0.001	<0.001	<0.001	0.01	<0.001
Moderator (Education)																	
No Qualifications		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Intermediate		1.00	0.97	1.03	0.75	0.87	0.18	1.03	0.77	1.02	0.92	0.85	0.34	0.95	0.80	1.24	0.10
Degree or higher		0.97	0.84	1.01	0.93	0.99	0.92	1.03	0.83	0.79	0.30	0.74	0.19	0.96	0.87	1.00	0.12
Predictor (Latent class)																	
Low-risk		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive		1.89	0.14	1.21	0.64	1.17	0.69	1.89	0.14	1.42	0.55	1.94	0.45	0.67	0.65	2.15	0.09
Low-risk yet heavy drinkers		1.68	0.27	0.57	0.36	0.89	0.81	1.68	0.27	0.41	0.17	1.59	0.65	0.11	0.33	2.32	0.12
Abstainers but inactive		1.89	0.10	1.21	0.61	1.42	0.31	1.89	0.10	2.07	0.16	2.94	0.19	0.96	0.95	2.31	0.04
Poor diet and inactive		2.40	0.05	0.96	0.92	1.97	0.08	2.40	0.05	1.27	0.70	1.26	0.80	1.06	0.94	2.60	0.04
Inactive, heavy drinkers		2.23	0.07	1.48	0.38	2.33	0.05	2.23	0.07	0.39	0.07	4.37	0.11	1.13	0.88	1.75	0.23
High-risk smokers		1.66	0.18	0.90	0.78	1.13	0.71	1.66	0.18	0.85	0.77	7.79	0.01	0.50	0.38	1.53	0.31
Moderation (Latent class x Education)																	
Low-risk (Ref.)	No Qualifications	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive	Intermediate	0.78	0.60	1.42	0.46	0.99	0.99	0.78	0.60	2.19	0.29	0.87	0.88	2.56	0.35	0.61	0.32
	Degree/higher	0.87	0.79	1.62	0.40	1.26	0.66	0.87	0.79	1.11	0.90	1.00	1.00	1.57	0.71	0.39	0.08
Low-risk yet heavy drinkers	Intermediate	0.60	0.33	1.42	0.61	1.29	0.61	0.60	0.33	1.56	0.63	0.89	0.91	19.77	0.19	0.53	0.26
	Degree/higher	0.81	0.69	1.74	0.43	1.12	0.84	0.81	0.69	1.17	0.87	0.60	0.67	15.66	0.24	0.31	0.05
Abstainers but inactive	Intermediate	0.80	0.62	1.78	0.18	0.91	0.82	0.80	0.62	1.90	0.35	0.79	0.79	1.76	0.51	0.54	0.16
	Degree/higher	1.58	0.40	2.58	0.10	1.37	0.56	1.58	0.40	1.71	0.49	0.30	0.21	2.41	0.40	0.65	0.44
Poor diet and inactive	Intermediate	0.41	0.07	1.23	0.67	0.35	0.02	0.41	0.07	1.01	0.99	1.26	0.82	1.92	0.48	0.49	0.15
	Degree/higher	0.35	0.09	1.65	0.47	0.76	0.62	0.35	0.09	1.73	0.54	0.47	0.51	1.99	0.55	0.30	0.07
Inactive, heavy drinkers	Intermediate	0.64	0.37	1.19	0.72	0.47	0.12	0.64	0.37	2.84	0.13	0.34	0.28	1.44	0.69	0.72	0.52
	Degree/higher	0.69	0.49	0.89	0.83	0.61	0.35	0.69	0.49	2.11	0.36	0.13	0.06	1.01	0.99	0.60	0.33
High-risk smokers	Intermediate	0.80	0.61	2.76	0.02	1.51	0.30	0.80	0.61	2.57	0.21	0.56	0.49	4.42	0.12	0.74	0.51
	Degree/higher	0.61	0.38	1.31	0.68	1.26	0.66	0.61	0.38	3.85	0.10	0.06	0.02	0.96	0.98	0.45	0.18

Bold values are statistically significant at p ≤0.005

Table 4. Tests of the moderating role of occupation in the relationship between risk behaviour latent class membership and disease outcomes

Disease outcomes		**Multimorbidity**		**Complex Multimorbidity**		**Circulatory disorders**		**Musculoskeletal disorders**		**Endocrine disorders**		**Respiratory disorders**		**Neoplasm**		**Eye disorders**	
		OR	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***
Intercept		0.03	<0.001	0.01	<0.001	0.12	<0.001	0.19	<0.001	0.19	0.05	0.03	0.001	0.00	<0.001	0.00	<0.001
Moderator (Occupation)																	
Routine/Manual		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Intermediate		0.85	0.10	0.90	0.30	1.03	0.73	1.01	0.96	0.82	0.21	1.02	0.93	1.19	0.38	0.10	0.13
Professional/managerial		0.87	0.14	0.87	0.18	0.88	0.17	0.97	0.80	1.01	0.95	0.67	0.02	1.12	0.57	0.14	0.64
Predictor (Latent class)																	
Low-risk		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive		1.54	0.14	1.81	0.05	1.09	0.75	0.99	0.96	2.62	0.14	1.97	0.28	1.42	0.63	0.14	0.02
Low-risk yet heavy drinkers		0.63	0.18	0.57	0.23	0.65	0.17	0.77	0.52	0.38	0.24	1.21	0.80	0.60	0.66	0.18	0.02
Abstainers but inactive		1.17	0.56	1.72	0.06	0.81	0.43	0.72	0.24	3.34	0.04	2.78	0.06	1.79	0.37	0.56	0.02
Poor diet and inactive		1.49	0.19	1.30	0.42	1.03	0.93	1.13	0.70	2.70	0.12	1.51	0.51	2.41	0.18	0.19	0.01
Inactive, heavy drinkers		1.51	0.18	1.71	0.08	1.28	0.38	1.30	0.39	1.03	0.96	1.76	0.36	2.14	0.24	0.18	0.03
High-risk smokers		1.15	0.62	1.84	0.04	0.98	0.95	1.19	0.53	1.92	0.29	**5.47***	**<0.001***	1.24	0.75	0.62	0.08
Moderation (Latent class x Education)																	
Low-risk (Ref.)	Routine/Manual	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive	Intermediate	1.13	0.79	0.86	0.75	1.20	0.68	0.99	0.97	0.88	0.88	1.32	0.72	0.45	0.40	0.79	0.25
	Professional	0.96	0.92	0.83	0.67	1.15	0.72	1.56	0.30	0.85	0.84	0.53	0.38	2.14	0.47	0.92	0.06
Low-risk yet heavy drinkers	Intermediate	2.78	0.03	1.72	0.38	2.90	0.02	1.07	0.90	3.40	0.25	2.77	0.26	1.99	0.59	0.03	0.04
	Professional	2.11	0.08	1.42	0.54	1.61	0.23	1.94	0.17	0.79	0.83	0.52	0.45	5.92	0.20	0.08	0.02
Abstainers but inactive	Intermediate	2.19	0.07	1.43	0.41	2.96	0.01	1.43	0.39	0.99	0.99	0.87	0.86	0.54	0.47	0.07	0.49
	Professional	1.72	0.17	1.12	0.79	2.46	0.02	1.68	0.20	1.16	0.84	0.40	0.19	1.43	0.72	0.17	0.16
Poor diet and inactive	Intermediate	0.50	0.13	0.70	0.47	0.99	0.98	0.46	0.08	0.13	0.07	0.36	0.30	0.19	0.09	0.13	0.02
	Professional	0.88	0.77	1.10	0.85	0.96	0.92	1.12	0.81	0.44	0.34	1.06	0.94	1.86	0.55	0.77	0.07
Inactive, heavy drinkers	Intermediate	1.35	0.49	0.94	0.90	1.51	0.33	0.64	0.33	0.81	0.82	1.17	0.85	0.15	0.05	0.49	0.26
	Professional	0.95	0.89	0.85	0.68	0.85	0.66	0.89	0.77	0.69	0.63	0.54	0.41	1.47	0.69	0.89	0.08
High-risk smokers	Intermediate	1.26	0.59	0.94	0.89	2.24	0.05	0.53	0.16	0.87	0.87	1.34	0.70	0.43	0.40	0.59	0.20
	Professional	1.30	0.51	0.71	0.46	1.62	0.20	1.09	0.83	0.88	0.87	0.28	0.11	3.29	0.27	0.51	0.03

Bold values are statistically significant at p ≤0.005

Age

Age had a positive main effect on the prevalence of multimorbidity, complex multimorbidity, circulatory disorders, neoplasms and eye disorders, such that older individuals were more likely to have these conditions (see Table 5). Latent class membership was also associated with complex multimorbidity. Specifically, we found that compared to the *Low-risk* cluster, the *Abstainer but Inactive* cluster (β = 7.32, p<0.001) was more likely to be diagnosed with complex multimorbidity. Finally, relationship between latent class membership and disease outcomes differed significantly as a function of age, such that older individuals in the *Abstainer but Inactive* cluster (β = -0.11, p<0.001) were less likely than those in the *Low-risk* cluster to have complex multimorbidity. However, age was not found to moderate any other relationships between cluster membership and disease outcomes.

Sex

Sex had a positive main effect on the prevalence of several disease outcomes (see Table 6). Compared to males, females had a higher likelihood of being diagnosed with multimorbidity (OR= 1.30, p<0.001), complex multimorbidity (OR =1.41, p< 0.001), musculoskeletal disorder (OR=2.05, p<0.001) and eye disorders (OR=1.53, p<0.001). On the other hand, females had lower odds of having endocrine disorders (0.59, p<0.001), and neoplasms (p=0.63, p<0.001), compared to males. Main effects of latent classes were significant for multimorbidity and respiratory disorders. Members in the *Low-risk yet inactive* cluster had nearly twice the odds of having multimorbidity (OR=2.08, p<0.001) compared to the *Low-risk* cluster. Similarly, *High-risk smokers* were nearly six times more likely to have respiratory disorders (OR=5.75, p<0.001), compared to the *Low-risk* cluster. However, we did not find differential effects of sex by latent class membership for any disease outcomes.

Table 5. Tests of the moderating role of age in the relationship between risk behaviour latent class membership and disease outcomes

Disease outcomes	**Multimorbidity**		**Complex Multimorbidity**		**Circulatory disorders**		**Musculoskeletal disorders**		**Endocrine disorders**		**Respiratory disorders**		**Neoplasm**		**Eye disorders**	
	β	***p-value***	***β***	***p-value***	***β***	***p-value***	***β***	***p-value***	***β***	***p-value***	***β***	***p-value***	***β***	***p-value***	***β***	***p-value***
Intercept	-2.34	0.30	-8.66	<0.001	-2.40	0.05	-3.08	0.01	-2.34	0.30	-4.20	0.06	-10.56	<0.001	-6.03	<0.001
Moderator (Age)																
Age (continuous)	**0.05***	**<0.001***	**0.05***	**<0.001***	**0.03***	**<0.001***	0.01	0.24	-0.01	0.17	<0.001	0.60	**0.05***	**<0.001***	**0.06***	**<0.001***
Predictor (Latent class)																
Low-risk	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive	0.30	0.92	4.37	0.01	1.33	0.43	1.95	0.27	0.30	0.92	2.23	0.39	3.72	0.31	2.61	0.07
Low-risk yet heavy drinkers	2.01	0.65	0.45	0.82	-2.76	0.13	1.34	0.45	2.01	0.65	-2.47	0.42	1.22	0.76	0.57	0.73
Abstainers but inactive	3.57	0.15	**7.32***	**<0.001***	0.92	0.58	2.09	0.16	3.57	0.15	4.05	0.13	6.81	0.07	2.76	0.04
Poor diet and inactive	-0.20	0.94	2.23	0.20	-1.30	0.47	-0.54	0.74	-0.20	0.94	0.58	0.84	5.20	0.17	1.14	0.49
Inactive, heavy drinkers	-2.02	0.51	4.43	0.01	1.63	0.30	2.43	0.14	-2.02	0.51	1.38	0.65	5.87	0.14	0.14	0.92
High-risk smokers	3.25	0.23	5.03	0.01	-0.18	0.92	1.52	0.37	3.25	0.23	-1.06	0.69	6.81	0.08	1.73	0.27
Moderation (Latent class x Age)																
Low-risk (Ref.)	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive	0.01	0.86	-0.06	0.02	-0.02	0.51	-0.03	0.31	0.01	0.86	-0.03	0.52	-0.05	0.32	-0.04	0.10
Low-risk yet heavy drinkers	-0.04	0.55	-0.01	0.74	0.05	0.12	-0.02	0.47	-0.04	0.55	0.05	0.36	-0.01	0.87	-0.01	0.80
Abstainers but inactive	-0.04	0.30	**-0.11***	**<0.001***	-0.01	0.63	-0.04	0.14	-0.04	0.30	-0.06	0.19	-0.10	0.07	-0.04	0.05
Poor diet and inactive	0.01	0.85	-0.03	0.21	0.02	0.47	0.01	0.79	0.01	0.85	-0.01	0.88	-0.07	0.20	-0.01	0.60
Inactive, heavy drinkers	0.03	0.55	-0.06	0.01	-0.02	0.38	-0.04	0.16	0.03	0.55	-0.02	0.73	-0.09	0.15	<0.001	0.97
High-risk smokers	-0.05	0.29	-0.07	0.01	0.01	0.76	-0.02	0.39	-0.05	0.29	0.04	0.35	-0.10	0.08	-0.03	0.29

Bold values are statistically significant at p≤0.005

Table 6. Tests of the moderating role of sex in the relationship between risk behaviour latent class membership and disease outcomes

Disease outcomes		**Multimorbidity**		**Complex Multimorbidity**		**Circulatory disorders**		**Musculoskeletal disorders**		**Endocrine disorders**		**Respiratory disorders**		**Neoplasm**		**Eye disorders**	
		OR	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***
Intercept		0.02	<0.001	0.01	<0.001	0.09	<0.001	0.18	<0.001	0.19	0.02	0.03	<0.001	<0.001	<0.001	0.01	<0.001
Moderator (Sex)																	
Male		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Female		**1.30***	**<0.001***	**1.41***	**<0.001***	1.06	0.48	**2.05***	**<0.001***	**0.59***	**<0.001***	0.95	0.70	**0.63***	**<0.001***	**1.53***	**<0.001***
Predictor (Latent class)																	
Low-risk		*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive		**2.08***	**<0.001***	1.80	0.04	1.62	0.07	1.11	0.70	3.13	0.01	3.08	0.01	1.25	0.64	1.67	0.04
Low-risk yet heavy drinkers		1.12	0.60	0.77	0.39	1.13	0.58	1.10	0.69	0.75	0.59	1.80	0.20	1.06	0.90	1.13	0.58
Abstainers but inactive		1.55	0.09	1.45	0.23	1.21	0.50	0.87	0.62	3.53	0.01	2.08	0.15	1.14	0.79	1.11	0.68
Poor diet and inactive		1.41	0.17	1.39	0.27	1.20	0.50	1.14	0.63	1.90	0.22	1.81	0.22	1.43	0.44	1.39	0.18
Inactive, heavy drinkers		1.77	0.01	1.58	0.08	1.45	0.08	1.05	0.83	1.05	0.91	1.72	0.21	1.27	0.58	1.32	0.17
High-risk smokers		1.16	0.54	1.43	0.22	1.41	0.15	0.77	0.32	1.77	0.24	**5.75***	**<0.001***	0.82	0.72	1.05	0.84
Moderation (Latent class x Sex)																	
Low-risk (Ref.)	Male	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Low-risk yet inactive	Female	0.63	0.17	0.87	0.71	0.62	0.16	1.07	0.84	0.55	0.33	0.35	0.07	1.18	0.84	0.65	0.17
Low-risk yet heavy drinkers	Female	1.31	0.45	1.03	0.95	0.90	0.75	0.89	0.77	0.20	0.21	0.57	0.41	3.10	0.18	1.07	0.85
Abstainers but inactive	Female	1.10	0.78	1.48	0.30	1.18	0.64	1.10	0.78	0.80	0.69	0.83	0.76	1.98	0.39	1.31	0.40
Poor diet and inactive	Female	0.73	0.37	0.76	0.50	0.77	0.46	0.70	0.34	0.56	0.41	0.47	0.24	1.69	0.52	0.94	0.86
Inactive, heavy drinkers	Female	0.72	0.32	0.96	0.92	0.78	0.42	1.04	0.91	0.56	0.45	0.82	0.73	1.02	0.98	0.90	0.71
High-risk smokers	Female	1.31	0.43	1.32	0.46	0.98	0.96	1.84	0.09	0.90	0.87	0.55	0.28	2.47	0.29	0.97	0.94

Bold values are statistically significant at p≤0.005

Discussion

The research reported in this paper investigated whether the relationship between clusters of health risk behaviours over time and disease outcomes is moderated by socio-demographic variables. Age, sex and wealth were each associated with multimorbidity, complex multimorbidity and, in some cases, other body system disorders. Education and occupation were not found to predict disease outcomes after controlling for other characteristics. We also found that compared to the *Low-risk* cluster, *High-risk smokers* had a higher likelihood of being diagnosed with respiratory disorders and complex multimorbidity, while the *Abstainer yet inactive* cluster was associated with a higher likelihood of complex multimorbidity. Notably, we found little evidence that socio-demographic characteristics moderate the relationship between behaviour latent classes and disease outcomes, with one exception where age moderated the relationship between latent classes and complex multimorbidity.

Effects of age and sex on disease outcomes

Our study found an overall positive association between age and the prevalence of multimorbidity, complex multimorbidity, circulatory disorders, neoplasms and eye disorders, such that older individuals were more likely to have these conditions. Age is a primary risk factor for multimorbidity (Violan et al., 2014) and numerous studies, including the present research, consistently demonstrate a positive correlation between age and a range of chronic diseases (Skou et al., 2022; Violan et al., 2014; Marengoni et al., 2011).

We also found that sex was associated with disease outcomes. Compared to males, females were more likely to have multimorbidity and complex multimorbidity. This finding corroborates existing evidence (Nguyen et al., 2019; Agur et al., 2016; Violan et al., 2014). Moreover, our study highlights a gender-specific predisposition for musculoskeletal disorders,

with females being twice as likely as males to suffer from these disorders, in line with global trends indicating a higher prevalence of osteoarthritis (Kloppenburg and Berenbaum, 2020) and rheumatoid arthritis (Finckh et al., 2022; Safiri et al., 2021). Conversely, males exhibited higher odds of endocrine disorders and neoplasms, with diabetes contributing significantly to the former. This sex disparity in endocrine disorders aligns with existing evidence documenting higher incidence and prevalence rates among men, particularly in older age groups in the UK (Zghebi et al., 2017; Holden et al., 2013). However, caution is warranted in interpreting the association between the male sex and neoplasms due to limited granularity on cancer types in the available data.

Effects of socio-economic status on disease outcomes

We studied three well-established dimensions of socio-economic status - wealth, education and occupation (Pathirana and Jackson, 2018). Lower household wealth emerged as a significant predictor for multimorbidity, complex multimorbidity, and endocrine disorders, consistent with a wealth of prior research indicating that material deprivation is consistently associated with higher multimorbidity and complex multimorbidity prevalence (Singer et al., 2019a; Álvarez-Gálvez et al., 2023). Equally, our results on endocrine disorders are consistent with previous studies using the ELSA study, which found particularly strong associations between lower wealth and incident and prevalent diabetes among older adults (Demakakos, Marmot and Steptoe, 2012; Nguyen et al., 2020).

However, in our study, occupation was not found to exert a main effect on disease outcomes after controlling for other measures of socio-economic status. This is broadly consistent with previous reviews examining the social determinants of multimorbidity, which found that

occupational levels were not definitively predictive of multimorbidity (Pathirana and Jackson, 2018; Álvarez-Gálvez et al., 2023). Similarly, we did not find education to be associated with multimorbidity after controlling for other measures of socio-economic status (Singer et al., 2019b). In part, these findings mirror a previous study examining latent classes of multimorbidity using the ELSA dataset, which found significantly strong associations for wealth and multimorbid classes (as compared to the relatively healthy class), while the associations for education were relatively attenuated (Nguyen et al., 2020). One explanation for such a pattern could be that in comparison to the measure of household wealth, which reflects a process of life-long accumulation, the indicators of education and occupation reflect periods of time from a more distant past. This might explain the stronger and more consistent effect of household wealth on disease outcomes, suggesting that it is a better indicator of an older person's status (McGovern and Nazroo, 2015; Adena and Myck, 2014). Equally, these findings highlight the nuanced role of socio-economic indicators in shaping disease outcomes among older populations.

Moderating effects of socio-demographic characteristics

Overall, we found little evidence to suggest that socio-demographic characteristics moderate the relationship between behaviour latent classes and disease outcomes. Overall, the absence of significant moderation effects, in spite of the main effects largely aligning with existing evidence, suggests one of two possibilities. Either (i) the present results accurately reflect that socio-demographic characteristics do not moderate the relationship between risk behaviour clusters and disease outcomes, or (ii) indicate that our analysis was statistically underpowered and, thus, unable to detect differences. The latter explanation – namely that the analysis was underpowered – seems more plausible due to the large number of moderating regression

models (n=40) and comparisons modelled on a relatively small sample size. With a larger population, it is possible that such a latent class moderation could not only be feasible but also informative in demonstrating how modifiable risk behaviour patterns interact with socio-demographic characteristics to influence disease outcomes.

An exception was the finding that age moderated the relationship between latent classes representing clusters of health behaviour over time and complex multimorbidity. Specifically, there was evidence that individuals in the *Abstainer but Inactive* cluster were less likely to have complex multimorbidity with increasing age compared to the *Low-risk* cluster, presenting an unexpected finding. In other words, younger individuals in the *Abstainer but Inactive* were more likely to have complex multimorbidity as compared to the *Low-risk* cluster. One plausible interpretation is that the influence of age on complex multimorbidity prevalence was mitigated by some other unaccounted risk factor, such as area-level socio-economic deprivation. Our prior study revealed that individuals characterised as *Abstainer but Inactive* were generally less affluent and had lower educational attainment compared to those in the *Low-risk* cluster, yet they experienced poorer health outcomes (Suhag, Webb and Holmes, 2024). Given that deprivation has been associated with earlier onset of multimorbidity, it is possible that despite controlling for individual-level socio-economic variables, more upstream social determinants of health could have played a role (Barnett et al., 2012). Nevertheless, caution is warranted in interpreting this finding without further analysis.

Limitations and Future Directions

A notable strength of our study was its consideration of clusters of behaviour over time through a multidimensional variable in a latent class moderation. Another strength of our study is that by incorporating longitudinal data, we were able to capture the dynamic nature of risk

behaviours and their impact on disease outcomes, addressing the criticism that epidemiological studies often only measure behaviour at one time point and thus provide a limited view of behaviour patterns and their changes over time. Finally, we used a Bonferroni correction to accommodate for the numerous comparisons, thus minimising potentially spurious significant results.

However, some limitations warrant consideration. The analysis is inherently complex since latent class moderation is a sophisticated technique that requires careful interpretation even at its simplest level. This approach necessarily involves simplifications which, at first, attempt to capture the dynamics of real-world behaviours. The model was then tasked with identifying subtle moderating effects within a multi-dimensional space defined by multiple socio-demographic characteristics and health outcomes—a considerable ask for even a well-powered statistical model. Due to these analytical challenges, the present conclusions must be viewed with the understanding that the breadth of comparisons and the size of our sample may have led to an underpowered statistical model. The wide confidence intervals in our moderation analysis further underscore the potential variability in our data and the possible effects of a small sample size.

Equally, we were unable to include all moderating effects in a single model, potentially missing some interactions between socio-demographic characteristics themselves. For instance, socio-demographic characteristics (e.g., education and income) may themselves interact and have important implications for both health behaviours and outcomes. For instance, an older adult with modest financial means but higher levels of education may navigate health management more effectively than economically comparable peers with less formal education, potentially influencing their proactive engagement with preventive health measures and ultimately

affecting their health outcomes (Perna et al., 2012). Acknowledging these complex interactions is important, even as modelling them is hard.

Additionally, the dietary measure in the ELSA relies primarily on weekly fruit and vegetable intake, which may not capture the full spectrum of dietary risks associated with chronic diseases as it fails to account for nutrient density and energy intake (Ocean, Howley and Ensor, 2019; Wirt and Collins, 2009). Using such a measure also misses issues like protein-energy malnutrition, which is prevalent among older adults and is linked to numerous health risks (del Carmen Alvarez-Nuncio and Ziegler, 2024; van der Pols-Vijlbrief et al., 2014). However, cohort studies like the Survey of Health, Ageing and Retirement in Europe (SHARE), the Lifelines Cohort Study, and NutriNet-Santé offer more comprehensive dietary assessments that can address these limitations. SHARE, for instance, includes a variety of protein-rich foods in its assessment, allowing us to evaluate adherence to a Mediterranean diet, which is known for its health benefits (Freisling et al., 2020; Kyprianidou et al., 2021). Similarly, the Lifelines study employs a detailed food frequency questionnaire that categorises foods based on their known health impacts (Vinke et al., 2018), while NutriNet-Santé uses a web-based 24-hour dietary record to track the intake of energy-dense, nutrient-poor ultra-processed foods, offering a more detailed view of dietary patterns and their health implications (Srour et al., 2019).

Finally, the relatively short time period (i.e. two years) between exposures (health risk behaviours) and disease outcomes, underscores the need for cautious interpretation regarding any potential causal relationships. This is because the possibility of a bidirectional relationship between risk behaviours and disease outcomes cannot be discounted.

Conclusion

In conclusion, the present research investigated whether the relationship between clusters of health behaviour over time and disease status was moderated by socio-demographic characteristics, finding little evidence for such moderation. It did, however, showcase how longitudinal health behaviour patterns can be modelled as a multidimensional predictor using latent class moderation, thus advancing methodological approaches for examining risk behaviours in relation to disease over time. Future research should explore multidimensional moderators to assess the interacting effects of multiple, diverse risk factors on chronic disease outcomes. Studies would benefit from using larger samples (e.g., UK Biobank, Understanding Society), comparing across international aging cohorts (e.g., using harmonised data from aging cohorts from the USA, England, Europe, Japan, Korea and China), and datasets offering more comprehensive measures for risk behaviours, particularly diet. Effective prevention and management of chronic diseases requires innovative approaches, and integrating recent statistical advancements will be crucial to designing effective behavioural interventions.

Supplementary Material

Table S1. Main effects of covariates with disease outcomes in moderation regression models for each socio-demographic characteristic (i.e. wealth, occupation, education, age and sex)

	Multimorbidity		Complex Multimorbidity		Circulatory disorders		Musculoskeletal disorders		Endocrine disorders		Respiratory disorders		Neoplasms		Eye disorders	
	p-value	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***	***p-value***	***OR***
Covariates (Wealth)																
Baseline disease	**<0.001***	**6.40***	**<0.001***	**6.83***	**<0.001***	**8.47***	**<0.001***	**17.96***	**<0.001***	**84.18***	**<0.001***	**61.81***	**<0.001***	**4.80***	**<0.001***	**2.51***
Female	**<0.001***	**1.30***	**<0.001***	**1.41***	0.55	1.05	**<0.001***	**2.04***	**<0.001***	**0.59***	0.78	0.96	**<0.001***	**0.61***	**<0.001***	**1.54***
Occupation - Intermediate	0.1	0.85	0.3	0.9	0.73	1.04	0.99	1	0.2	0.81	0.98	1	0.34	1.21	0.13	0.86
Occupation -	0.87	0.2	0.87	0.18	0.88	0.83	0.98	0.99	1	0.02	0.67	0.56	1.13	0.72	0.97	
Education - Intermediate	0.92	1.01	0.73	1.04	0.22	0.88	0.74	1.04	0.89	1.02	0.34	0.86	0.86	0.97	0.04	1.23
Education - Degree or higher	0.9	0.98	0.96	1.01	1	1	0.8	1.04	0.31	0.79	0.2	0.74	0.82	0.94	0.09	1.24
Age (Cont.)	**<0.001***	**1.05***	**<0.001***	**1.05***	**<0.001***	**1.03***	0.21	1.01	0.17	0.99	0.63	1	**<0.001***	**1.05**	**<0.001***	**1.06***
Covariates(Education)																
Baseline disease	**<0.001**	**6.40**	**<0.001**	**6.94**	**<0.001**	**8.39**	**<0.001**	**6.40**	**<0.001**	**80.88**	**<0.001**	**62.12**	**<0.001**	**4.93**	**<0.001**	**2.53**
Female	**<0.001**	**1.29**	**<0.001**	**1.40**	0.60	1.04	**<0.001**	**1.29**	**<0.001**	**0.59**	0.74	0.96	**<0.001**	**0.62**	**<0.001**	**1.53**
Occupation - Intermediate	0.12	0.85	0.32	0.90	0.68	1.04	0.12	0.85	0.22	0.82	0.88	1.03	0.36	1.20	0.13	0.87
Occupation -	0.14	0.86	0.17	0.86	0.17	0.87	0.14	0.86	0.93	1.02	0.02	0.67	0.55	1.13	0.59	0.95
Age	**<0.001**	**1.05**	**<0.001**	**1.05**	**<0.001**	**1.03**	**<0.001**	**1.05**	0.17	0.99	0.59	1.01	**<0.001**	**1.05**	**<0.001**	**1.06**
Wealth - 2nd Tertile	0.08	0.83	0.53	0.94	0.23	0.88	0.08	0.83	**<0.001**	**0.59**	0.88	0.98	0.77	1.06	0.24	1.12
Wealth - 3rd Tertile	**<0.001**	**0.67**	**0.01**	**0.73**	0.12	0.84	**<0.001**	**0.67**	**<0.001**	**0.60**	0.66	1.08	0.14	0.74	0.97	1.00
Covariates (Occupation)																
Baseline disease	**<0.001**	**6.48**	**<0.001**	**6.88**	**<0.001**	**8.46**	**<0.001**	**18.07**	**<0.001**	**86.40**	**<0.001**	**63.31**	**<0.001**	**5.13**	**<0.001**	**2.53**
Female	**<0.001**	**1.30**	**<0.001**	**1.41**	0.52	1.05	**<0.001**	**2.05**	**<0.001**	**0.59**	0.71	0.95	0.01	0.64	**<0.001**	**1.53**
Education - Intermediate	0.98	1.00	0.71	1.04	0.14	0.86	0.75	1.04	0.90	1.02	0.39	0.87	0.84	0.96	0.03	1.24
Education - Degree or higher	0.91	0.99	0.88	1.02	0.96	1.01	0.76	1.05	0.41	0.82	0.18	0.73	0.97	0.99	0.11	1.23
Age	**<0.001**	**1.05**	**<0.001**	**1.05**	**<0.001**	**1.03**	0.27	1.01	0.16	0.99	0.65	1.00	**<0.001**	**1.05**	**<0.001**	**1.06**
Wealth - 2nd Tertile	0.05	0.81	0.42	0.92	0.14	0.86	0.78	0.97	**<0.001**	**0.59**	0.89	0.98	0.88	1.03	0.26	1.11
Wealth - 3rd Tertile	**<0.001**	**0.66**	**<0.001**	**0.71**	0.10	0.83	0.06	0.80	**0.01**	**0.61**	0.64	1.08	0.13	0.73	0.90	0.99
Covariate (Age)																
Baseline disease	**1.87***	**<0.001***	**1.96***	**<0.001***	**2.13***	**<0.001***	**2.89***	**<0.001***	**4.41****	**<0.001***	**4.14***	**<0.001***	**1.65***	**<0.001***	**0.93***	**<0.001***
Female	0.25	0.01	**0.28***	**0.01***	0	1	**0.71***	**<0.001***	**-0.60***	**<0.001***	-0.13	0.4	-0.49	0.01	**0.37***	**<0.001***

Occupation - Intermediate		-0.17	0.09	-0.1	0.33	0.04	0.71	0.01	0.9	-0.19	0.23	0.03	0.88	0.18	0.37	-0.14	0.15
Occupation	-	-0.15	0.14	-0.15	0.18	-0.13	0.18	-0.03	0.8	0.01	0.97	-0.38	0.03	0.12	0.55	-0.04	0.65
Education - Intermediate		0	0.99	0.01	0.91	-0.14	0.18	0.03	0.8	0	0.99	-0.16	0.34	-0.08	0.68	0.2	0.04
Education - Degree or higher		-0.03	0.83	0	0.99	-0.02	0.91	0.03	0.83	-0.26	0.27	-0.32	0.16	-0.07	0.82	0.2	0.11
Wealth - 2nd Tertile		-0.2	0.05	-0.08	0.46	-0.15	0.14	-0.02	0.85	**-0.53***	**<0.001***	-0.04	0.82	0.04	0.85	0.1	0.29
Covariates (Sex)																	
Baseline disease		**<0.001***	**6.39***	**<0.001***	**6.86***	**<0.001***	**8.35***	**<0.001***	**17.74****	**<0.001***	**79.84****	**<0.001****	**60.52****	**<0.001***	**4.95***	**<0.001***	**2.50***
Occupation - Intermediate		0.11	0.85	0.32	0.9	0.72	1.04	0.94	1.01	0.24	0.83	0.94	1.01	0.35	1.2	0.14	0.87
Occupation	-	0.14	0.86	0.17	0.86	0.17	0.87	0.79	0.97	0.95	1.01	0.02	0.67	0.53	1.14	0.65	0.96
Education - Intermediate		0.99	1	0.79	1.03	0.17	0.87	0.76	1.03	0.91	1.02	0.32	0.85	0.77	0.94	0.03	1.23
Education - Degree or higher		0.81	0.97	0.97	1.01	0.89	0.98	0.8	1.04	0.3	0.79	0.19	0.74	0.83	0.94	0.11	1.23
Age		**<0.001***	**1.05***	**<0.001***	**1.05***	**<0.001****	**1.03****	0.24	1.01	0.16	0.99	0.62	1	**<0.001***	**1.05***	**<0.001***	**1.06***
Wealth - 2nd Tertile		0.04	0.81	0.42	0.92	0.14	0.86	0.82	0.98	**<0.001****	**0.59***	0.84	0.97	0.85	1.04	0.3	1.1
Wealth - 3rd Tertile		**<0.001***	**0.66***	**<0.001***	**0.71***	0.08	0.83	0.07	0.81	**<0.001***	**0.60***	0.72	1.06	0.13	0.73	0.87	0.98

***Bold values** are statistically significant the significance level (p=0.005)

References

Adena, M. & Myck, M. 2014. Poverty and transitions in health in later life. *Social science & medicine,* 116**,** 202-210.

Agur, K., McLean, G., Hunt, K., Guthrie, B. & Mercer, S. W. 2016. How does sex influence multimorbidity? Secondary analysis of a large nationally representative dataset. *International journal of environmental research and public health,* 13**,** 391.

Álvarez-Gálvez, J., Ortega-Martín, E., Carretero-Bravo, J., Pérez-Muñoz, C., Suárez-Lledó, V. & Ramos-Fiol, B. 2023. Social determinants of multimorbidity patterns: A systematic review. *Frontiers in Public Health,* 11**,** 1081518.

Artaud, F., Sabia, S., Dugravot, A., Kivimaki, M., Singh-Manoux, A. & Elbaz, A. 2016. Trajectories of unhealthy behaviors in midlife and risk of disability at older ages in the Whitehall II Cohort Study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences,* 71**,** 1500-1506.

Asparouhov, T. & Muthén, B. 2014. Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus web notes,* 21**,** 1-22.

Banks, J., Phelps, A., Oskala, A., Steptoe, A., Blake, M., Oldfield, Z., Marmot, M., Clemens, S., Rogers, N., Nazroo, J. 2021. English Longitudinal Study of Ageing: Waves 0-9, 1998-2019.

Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S. & Guthrie, B. 2012. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet,* 380**,** 37-43.

Bricca, A., Jäger, M., Johnston, M., Zangger, G., Harris, L. K., Midtgaard, J. & Skou, S. T. 2023. Effect of in-person delivered behavioural interventions in people with multimorbidity: systematic review and meta-analysis. *International Journal of Behavioral Medicine,* 30**,** 167-189.

Cano, M. Á., de Dios, M. A., Correa-Fernández, V., Childress, S., Abrams, J. L. & Roncancio, A. M. 2017. Depressive symptom domains and alcohol use severity among Hispanic emerging adults: Examining moderating effects of gender. *Addictive behaviors,* 72**,** 72-78.

del Carmen Alvarez-Nuncio, M. & Ziegler, T. R. 2024. Micronutrient status and protein-energy malnutrition in free-living older adults: a current perspective. *Current opinion in gastroenterology,* 40**,** 99-105.

Demakakos, P., Marmot, M. & Steptoe, A. 2012. Socio-economic position and the incidence of type 2 diabetes: the ELSA study. *European journal of epidemiology,* 27**,** 367-378.

Desmond, C., Seeley, J., Groenewald, C., Ngwenya, N., Rich, K. & Barnett, T. 2019. Interpreting social determinants: Emergent properties and adolescent risk behaviour. *PLoS One,* 14**,** e0226241.

Dhalwani, N. N., O'Donovan, G., Zaccardi, F., Hamer, M., Yates, T., Davies, M. & Khunti, K. 2016. Long terms trends of multimorbidity and association with physical activity in older English population. *International Journal of Behavioral Nutrition and Physical Activity,* 13**,** 1-9.

Diez Roux, A. V. 2004. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiologic reviews,* 26**,** 104-111.

Finckh, A., Gilbert, B., Hodkinson, B., Bae, S.-C., Thomas, R., Deane, K. D., Alpizar-Rodriguez, D. & Lauper, K. 2022. Global epidemiology of rheumatoid arthritis. *Nature Reviews Rheumatology,* 18**,** 591-602.

Fraser, S. D., Roderick, P. J., Aitken, G., Roth, M., Mindell, J. S., Moon, G. & O'Donoghue, D. 2014. Chronic kidney disease, albuminuria and socio-economic status in the Health Surveys for England 2009 and 2010. *Journal of Public Health,* 36**,** 577-586.

Freisling, H., Viallon, V., Lennon, H., Bagnardi, V., Ricci, C., Butterworth, A. S., Sweeting, M., Muller, D., Romieu, I. & Bazelle, P. 2020. Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. *BMC medicine,* 18**,** 1-11.

Hamer, M., de Oliveira, C. & Demakakos, P. 2014. Non-exercise physical activity and survival: English longitudinal study of ageing. *American journal of preventive medicine,* 47**,** 452-460.

Harrison, C., Britt, H., Miller, G. & Henderson, J. 2014. Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. *BMJ open,* 4**,** e004694.

He, L., Biddle, S. J., Lee, J. T., Duolikun, N., Zhang, L., Wang, Z. & Zhao, Y. 2021. The prevalence of multimorbidity and its association with physical activity and sleep duration in middle aged and elderly adults: a longitudinal analysis from China. *International Journal of Behavioral Nutrition and Physical Activity,* 18**,** 77.

Head, A., Fleming, K., Kypridemos, C., Schofield, P. & O'Flaherty, M. 2020. Dynamics of multimorbidity in England between 2004 and 2019: a descriptive epidemiology study. *European Journal of Public Health,* 30**,** ckaa165. 818.

Head, A., Fleming, K., Kypridemos, C., Schofield, P., Pearson-Stuttard, J. & O'Flaherty, M. 2021. Inequalities in incident and prevalent multimorbidity in England, 2004–19: a population-based, descriptive study. *The Lancet Healthy Longevity,* 2**,** e489-e497.

Health, D. o. 2016. Alcohol Guidelines Review – Report from the Guidelines Development Group to the UK Chief Medical Officer. .

Hoekstra, T. 2013. Applied latent class models for epidemiology.

Holden, S., Barnett, A. H., Peters, J. R., Jenkins-Jones, S., Poole, C. D., Morgan, C. L. & Currie, C. J. 2013. The incidence of type 2 diabetes in the United Kingdom from 1991 to 2010. *Diabetes, Obesity and Metabolism,* 15**,** 844-852.

Hutchesson, M. J., Duncan, M. J., Oftedal, S., Ashton, L. M., Oldmeadow, C., Kay-Lambkin, F. & Whatnall, M. C. 2021. Latent class analysis of multiple health risk behaviors among Australian university students and associations with psychological distress. *Nutrients,* 13**,** 425.

Irizar, P., Gage, S. H., Fallon, V. & Goodwin, L. 2022. A latent class analysis of health risk behaviours in the UK Police Service and their associations with mental health and job strain. *BMC psychiatry,* 22**,** 426.

Kershaw, K. N., Mezuk, B., Abdou, C. M., Rafferty, J. A. & Jackson, J. S. 2010. Socio-economic position, health behaviors, and C-reactive protein: a moderated-mediation analysis. *Health Psychology,* 29**,** 307.

Kingston, A., Comas-Herrera, A. & Jagger, C. 2018. Forecasting the care needs of the older population in England over the next 20 years: estimates from the Population Ageing and Care Simulation (PACSim) modelling study. *The Lancet Public Health,* 3**,** e447-e455.

Kino, S., Bernabé, E. & Sabbah, W. 2017. Socio-economic inequality in clusters of health-related behaviours in Europe: latent class analysis of a cross-sectional European survey. *BMC Public Health,* 17**,** 1-8.

Kloppenburg, M. & Berenbaum, F. 2020. Osteoarthritis year in review 2019: epidemiology and therapy. *Osteoarthritis and cartilage,* 28**,** 242-248.

Kyprianidou, M., Panagiotakos, D., Faka, A., Kambanaros, M., Makris, K. C. & Christophi, C. A. 2021. Adherence to the Mediterranean diet in Cyprus and its relationship to multimorbidity: An epidemiological study. *Public Health Nutrition,* 24**,** 4546-4555.

Leventhal, A. M., Huh, J. & Dunton, G. F. 2014. Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis. *Perspectives in Public Health,* 134**,** 331-338.

Luk, J. W. & Tsoh, J. Y. 2010. Moderation of gender on smoking and depression in Chinese Americans. *Addictive Behaviors,* 35**,** 1040-1043.

Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B. & Fratiglioni, L. 2011. Aging with multimorbidity: a systematic review of the literature. *Ageing research reviews,* 10**,** 430-439.

McGovern, P. & Nazroo, J. Y. 2015. Patterns and causes of health inequalities in later life: a Bourdieusian approach. *Sociology of Health & Illness,* 37**,** 143-160.

McKnight, A. & Elias, P. 2003. Earnings, unemployment and the NS-SEC.

Mercer, S., Furler, J., Moffat, K., Fischbacher-Smith, D. & Sanci, L. 2016. *Multimorbidity: technical series on safer primary care*, World Health Organization.

Mishra, V. K., Srivastava, S. & Murthy, P. 2021. Population attributable risk for multimorbidity among adult women in India: Do smoking tobacco, chewing tobacco and consuming alcohol make a difference? *PloS one,* 16**,** e0259578.

Muthén, B. & Muthén, L. 2017. Mplus. *Handbook of item response theory.* Chapman and Hall/CRC.

Myint, P. K., Smith, R. D., Luben, R. N., Surtees, P. G., Wainwright, N. W., Wareham, N. J. & Khaw, K.-T. 2011. Lifestyle behaviours and quality-adjusted life years in middle and older age. *Age and ageing,* 40**,** 589-595.

Nguyen, H., Chua, K.-C., Dregan, A., Vitoratou, S., Bayes-Marin, I., Olaya, B. & Prina, A. M. 2020. Factors associated with multimorbidity patterns in older adults in England: findings from the English Longitudinal Study of Aging (ELSA). *Journal of aging and health,* 32**,** 1120-1132.

Nguyen, H., Manolova, G., Daskalopoulou, C., Vitoratou, S., Prince, M. & Prina, A. M. 2019. Prevalence of multimorbidity in community settings: A systematic review and meta-analysis of observational studies. *Journal of comorbidity,* 9**,** 2235042X19870934.

Noble, N., Paul, C., Turon, H. & Oldmeadow, C. 2015. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. *Preventive medicine,* 81**,** 16-41.

Nylund-Gibson, K. & Choi, A. Y. 2018. Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science,* 4**,** 440.

Ocean, N., Howley, P. & Ensor, J. 2019. Lettuce be happy: A longitudinal UK study on the relationship between fruit and vegetable consumption and well-being. *Social Science & Medicine,* 222**,** 335-345.

Pathirana, T. I. & Jackson, C. A. 2018. Socio-economic status and multimorbidity: a systematic review and meta-analysis. *Australian and New Zealand journal of public health,* 42**,** 186-194.

Perna, L., Mielck, A., Lacruz, M. E., Emeny, R. T., Holle, R., Breitfelder, A. & Ladwig, K. H. 2012. Socio-economic position, resilience, and health behaviour among elderly people. *International journal of public health,* 57**,** 341-349.

Prochaska, J. J., Nigg, C. R., Spring, B., Velicer, W. F. & Prochaska, J. O. 2010. The benefits and challenges of multiple health behavior change in research and in practice. *Preventive medicine,* 50**,** 26-29.

Safiri, S., Kolahi, A. A., Cross, M., Hill, C., Smith, E., Carson-Chahhoud, K., Mansournia, M. A., Almasi-Hashiani, A., Ashrafi-Asgarabad, A. & Kaufman, J. 2021. Prevalence, deaths, and disability-adjusted life years due to musculoskeletal disorders for 195 countries and territories 1990–2017. *Arthritis & rheumatology,* 73**,** 702-714.

Singer, L., Green, M., Rowe, F., Ben-Shlomo, Y., Kulu, H. & Morrissey, K. 2019a. Trends in multimorbidity, complex multimorbidity and multiple functional limitations in the ageing population of England, 2002–2015. *Journal of comorbidity,* 9**,** 2235042X19872030.

Singer, L., Green, M., Rowe, F., Ben-Shlomo, Y. & Morrissey, K. 2019b. Social determinants of multimorbidity and multiple functional limitations among the ageing population of England, 2002–2015. *SSM-population health,* 8**,** 100413.

Skou, S. T., Mair, F. S., Fortin, M., Guthrie, B., Nunes, B. P., Miranda, J. J., Boyd, C. M., Pati, S., Mtenga, S. & Smith, S. M. 2022. Multimorbidity. *Nature Reviews Disease Primers,* 8**,** 48.

Soley-Bori, M., Ashworth, M., Bisquera, A., Dodhia, H., Lynch, R., Wang, Y. & Fox-Rushby, J. 2021. Impact of multimorbidity on healthcare costs and utilisation: a systematic review of the UK literature. *British Journal of General Practice,* 71**,** e39-e46.

Srour, B., Fezeu, L. K., Kesse-Guyot, E., Allès, B., Méjean, C., Andrianasolo, R. M., Chazelas, E., Deschasaux, M., Hercberg, S. & Galan, P. 2019. Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Santé). *bmj,* 365.

Suhag, A., Webb, T. L. & Holmes, J. 2024. Longitudinal clustering of health behaviours and their association with multimorbidity in older adults in England: A latent class analysis. *Plos one,* 19**,** e0297422.

Tegegne, T. K., Islam, S. M. S. & Maddison, R. 2022. Effects of lifestyle risk behaviour clustering on cardiovascular disease among UK adults: latent class analysis with distal outcomes. *Scientific Reports,* 12**,** 17349.

Tsimpida, D., Kontopantelis, E., Ashcroft, D. M. & Panagioti, M. 2022. The dynamic relationship between hearing loss, quality of life, socio-economic position and depression and the impact of hearing aids: answers from the English Longitudinal Study of Ageing (ELSA). *Social psychiatry and psychiatric epidemiology,* 57**,** 353-362.

van der Pols-Vijlbrief, R., Wijnhoven, H. A., Schaap, L. A., Terwee, C. B. & Visser, M. 2014. Determinants of protein–energy malnutrition in community-dwelling older adults: a systematic review of observational studies. *Ageing research reviews,* 18**,** 112-131.

Vinke, P. C., Corpeleijn, E., Dekker, L. H., Jacobs Jr, D. R., Navis, G. & Kromhout, D. 2018. Development of the food-based Lifelines Diet Score (LLDS) and its application in 129,369 Lifelines participants. *European journal of clinical nutrition,* 72**,** 1111-1119.

Violan, C., Foguet-Boreu, Q., Flores-Mateo, G., Salisbury, C., Blom, J., Freitag, M., Glynn, L., Muth, C. & Valderas, J. M. 2014. Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. *PloS one,* 9**,** e102149.

Willadsen, T. G., Bebe, A., Køster-Rasmussen, R., Jarbøl, D. E., Guassora, A. D., Waldorff, F. B., Reventlow, S. & Olivarius, N. d. F. 2016. The role of diseases, risk factors and symptoms in the definition of multimorbidity–a systematic review. *Scandinavian journal of primary health care,* 34**,** 112-121.

Wirt, A. & Collins, C. E. 2009. Diet quality–what is it and does it matter? *Public health nutrition,* 12**,** 2473-2492.

Zghebi, S. S., Steinke, D. T., Carr, M. J., Rutter, M. K., Emsley, R. A. & Ashcroft, D. M. 2017. Examining trends in type 2 diabetes incidence, prevalence and mortality in the UK between 2004 and 2014. *Diabetes, Obesity and Metabolism,* 19**,** 1537-1545.

Zhu, Y., Steele, F. & Moustaki, I. 2017. A general 3-step maximum likelihood approach to estimate the effects of multiple latent categorical variables on a distal outcome. *Structural Equation Modeling: A Multidisciplinary Journal,* 24**,** 643-656.
Zwolinsky, S., Raine, G. & Robertson, S. 2016. Prevalence, co-occurrence and clustering of lifestyle risk factors among UK men. *Journal of Men's Health,* 12.

7. General discussion and conclusion

7.1 Overview of the thesis

The overall aim of this thesis was to advance our understanding of how health-risk behaviours (i.e. smoking, poor nutrition, alcohol consumption, and physical inactivity) cluster over time and whether and how the identified clusters are prospectively associated with the prevalence of multimorbidity and a range of chronic disease outcomes. To do this, I used data from two large, ongoing nationally-representative panel studies in the UK. The first of older adults (aged 50 and over) and the second of adults (aged 16 and over). I first examined how SNAP risk behaviours (i.e. smoking, poor nutrition, alcohol consumption, and physical inactivity) cluster together and how behaviours within these clusters change as people age. Then, I examined the sociodemographic profiles and prospective disease outcomes associated with each of the identified clusters.

These aims were achieved across the first two research studies in this thesis (detailed in Chapters 4 and 5), whose findings complement, reinforce, and build upon each other to further explain the relationship(s) between risk behaviours, including i) how and why they may cluster together, ii) how they are associated with sociodemographic characteristics like age, sex, education, occupation and income/wealth, and iii) how they affect multimorbidity and a range of other disorders. Chapter 6 investigated the possibility of sociodemographic characteristics moderating the relationship between risk behaviour clusters and disease outcomes in older adults, using the clusters identified in Chapter 4.

This final chapter will integrate findings from the three studies and contextualise them against the wider literature. Following this, I will discuss the overarching implications of the research for future research and public health policy, identifying areas for further investigation and potential avenues for intervention. This will be followed by a critical reflection on the study design, acknowledging both its strengths and limitations, and offering insights into the methodological considerations that shaped our findings and could guide future studies. The final section will draw conclusions with respect to the contribution of this work in the wider research context.

7.1.1 Clustering of risk behaviours and behavioural patterns within clusters over time

The first aim of the studies presented in Chapters 4 and 5 was to investigate whether and how SNAP behaviours cluster over time in a nationally-representative sample of older adults (aged 50 years or older) and a general adult sample (aged 16 years or older), respectively. Previous studies using clustering techniques to analyse multiple behaviours in older adults or the general adult population have typically been cross-sectional, and so cannot test whether patterns of risk behaviour within clusters change over time and if these changes affect long-term health (Schneider et al., 2009; Liao et al., 2019; Griffin et al., 2014). To our knowledge, only two studies have used longitudinal clustering to study a subset of SNAP behaviours in older adults. They used data from China and Taiwan, respectively (Hsu et al., 2013; Feng et al., 2022). No studies have examined these relationships in the context of the UK population. Thus, the studies in this thesis break fresh ground in being the first of their kind to examine the longitudinal clustering of multiple health-risk behaviours across multiple time-points in the UK context.

The study on older adults using the ELSA data (see Chapter 4 for details) identified seven clusters: (i) *Low-risk* (13.4%), (ii) *Low-risk yet inactive (*16.8%), (iii) *Low-risk yet heavy drinkers* (11.4%), (iv) *Abstainer yet inactive* (20.0%), (v) *Poor diet and inactive* (12.9%), (vi) *Inactive, heavy drinkers* (14.5%), and (vii) *High-risk smokers* (10.9%). The study in Chapter 5 replicated the analysis conducted in Chapter 4 across a broader age range to examine SNAP behaviour patterns at earlier stages of the life course and used data from a nationally-representative sample of adults (aged 16 years and over) drawn from the UKHLS panel study. Seven clusters representing distinct profiles of health behaviours were identified: *Overall low-risk* (20%), *Insufficiently active* (18.3%), *Poor diet, Insufficiently active* (23.2%), *Hazardous and harmful drinkers* (10.7%), *Hazardous drinkers, Insufficiently active, and Poor diet* (14%), *Smokers and drinkers* (5.1%), and *Smokers* (8.7%).

There were some notable similarities in the clusters between our studies using the ELSA and UKHLS datasets. First, the four core clusters mentioned above, namely — an overall low-risk cluster, a cluster with insufficient physical activity as the only risky behaviour, a cluster characterised by poor diet and insufficient physical activity, and a cluster of current smokers with multiple risky behaviours — were also represented in our study using the ELSA study. However, the ELSA study had only one cluster with current smokers, while the UKHLS had two. This finding is in keeping with the observations that the odds of current smoking progressively decrease with increases in age beyond 25 years, thus explaining the smaller proportion of current smokers in the overall sample of older adults examined (Opazo Breton et al., 2022).

The patterns of behaviour observed within the clusters were largely stable over time in both studies. The finding that behavioural patterns are relatively stable over time in both studies was consistent with prior evidence suggesting that risk behaviours are relatively stable during mid-adulthood (Mawditt et al., 2016; Benzies, Wångby and Bergman, 2008) and older age (Botoseneanu and Liang, 2012), in contrast to adolescence and young adulthood (Daw, Margolis and Wright, 2017; Fromme, Corbin and Kruse, 2008). The stability in risk behaviour patterns within clusters suggests that clustering in the general adult samples and older adults can be accurately captured by cross-sectional studies. More importantly, the stability of behaviour patterns during mid-life implies that these behaviours could be sustained across adulthood, which is an important period in the life course when individuals play a central role in supporting younger and older generations and make a valuable contribution to the workforce and society (Lachman, Teshale and Agrigoroaei, 2015). Given that prolonged patterns of risky behaviours among adults may not only increase the risk of disease but also adversely impact the next generation through their role as parents (Mawditt et al., 2016), these findings underscore the importance of addressing risk behaviours early in the life course to prevent negative health outcomes.

There were, however, occasional exceptions to the dominant pattern of similarity. Importantly, these exceptions were comparable across the two studies, indicating a level of consistency and reliability in the observed behavioural dynamics. For instance, in the study on the general adult sample in UKHLS, the proportion of people consuming alcohol at low-risk levels (compared to hazardous levels) rose across waves in the clusters characterised by low levels of alcohol consumption (i.e. *Overall low-risk and Poor diet, Insufficiently active* clusters*).* This trend was mirrored in behaviour patterns within the clusters of older adults in ELSA, where the proportion of alcohol abstainers gradually increased in clusters characterised by moderate or

no alcohol consumption (i.e. the clusters labelled *Low-risk* and *Abstainer yet inactive*). Similarly, the proportion of current smokers steadily declined in the *High-risk smokers* cluster of older adults in ELSA and the proportion of current smokers gradually rose over time to then decline in the last wave in the *Smokers and Drinkers* cluster. In the UKHLS, the proportion of smokers consistently fell over time in the *Smokers* cluster, in line with the reduction in current smoking rates observed in adults of working age or older in the UK (Opazo Breton et al., 2022). These exceptions underscore a notable trend: where changes across time are observed, health-risk behaviours tend to improve with age.

7.1.2 Sociodemographic profiles of risk behaviour clusters

The clusters identified in each study also had distinct socio-demographic profiles, with notable similarities between the sociodemographic characteristics of clusters in the two samples. Consistent with patterns of alcohol consumption in the UK (Meier et al., 2021), the two clusters of heavy drinkers in the older sample, and the three clusters of hazardous and harmful drinkers in the general adult sample, were predominantly male. In the older sample, *High-risk smokers* were younger on average and, in contrast to previous research, were more likely to be women (Barta, Powell and Wisnivesky, 2019). These *High-risk smokers* bore a marked resemblance to the *Smokers* cluster in the general adult sample, who were also more likely to be women and of working age. Importantly, both clusters also exhibited socio-economic disadvantage.

There were additional similarities in the sociodemographic characteristics of the clusters found in the two samples. For instance, the *Abstainer yet inactive* cluster in the older sample and the *Overall low-risk* cluster in the general adult sample had similar sociodemographic profiles in

that both clusters had a female majority, were likely to be older, were least likely to belong to the highest quintile of wealth and most likely to hold no educational qualifications. Conversely, the socio-demographic and behavioural profile of the *Poor diet, Insufficiently active* cluster in UKHLS appears to closely resemble that of the 'Healthy' or 'Safe' clusters identified in prior studies - including the *Low-risk yet Inactive* cluster identified in previous results using the ELSA study. These clusters included primarily older women from advantaged socio-economic groups characterised by higher levels of education and wealth (Conry et al., 2011; French, Rosenberg and Knuiman, 2008). Overall, the older female-majority and behavioural profiles of these two sets of clusters are consistent with previous findings on risk behavioural clustering (Schneider et al., 2009; Griffin et al., 2014; Liao et al., 2019; Hsu et al., 2013).

However, there were some differences in the sociodemographic characteristics of the clusters found in the two samples. First, there was an additional cluster in UKHLS, namely, *Smokers and Drinkers*, that was characterised by a high likelihood of current smoking. The *Smokers and Drinkers* cluster had the highest proportion of younger individuals (aged 16-34) of all clusters and a two-thirds male majority, consistent with previous research indicating a clustering of alcohol consumption and smoking among young adult males (Noble et al., 2015). Similarly, although we could not test for ethnic differences in the older sample due to relatively few ethnic minority participants in the ELSA study, there were differences in the ethnic distribution of clusters in the general adult sample. Specifically, the minority groups examined in the UKHLS (i.e. Indian, Pakistani, Bangladeshi, Black African and Black Caribbean) were more likely to fall in the *Overall low-risk* cluster, consistent with other UK studies that have pointed to healthier behaviours among minority ethnic groups (Graham et al., 2016; Lawder et al., 2010).

In sum, there was a broad consistency across the two samples in the associations between clusters and sociodemographic variables. That is, the clusters with similar profiles of SNAP behaviour patterns across the two samples also tended to resemble each other in the sociodemographic characteristics of the participants that comprised those clusters. Thus, the observation that two different datasets associate individuals with similar socio-demographic attributes with the same broad patterns of behaviour, suggests that these clusters reflect reasonably robust targets for intervention. This suggests that identifying individuals with similar sociodemographic profiles may help predict risk behaviour patterns without directly measuring behaviour. Leveraging this insight, interventions can be efficiently targeted toward groups more likely to exhibit risky patterns of behaviour.

7.1.3 Associations between clusters of risk behaviour and disease outcomes

The most important contribution of this thesis was to examine whether and how longitudinal clusters of risk behaviours are associated with disease outcomes, specifically multimorbidity. Here too, distinctive similarities in the results from the two samples were noted.

In both samples, some associations between clusters and disease outcomes were easier to anticipate than others. For instance, in the older sample, *High-risk smokers* were most likely to have respiratory disorders as might be expected on the basis of extensive evidence that smoking damages lung function (Jayes et al., 2016). Similarly, in the general adult sample, both *Smokers* and *Smokers and Drinkers* had a similar prevalence of respiratory disorders, although *Smokers* had a higher prevalence of respiratory disorders than the *Insufficiently Active* cluster,

aligning with established research implicating cigarette smoking as a key risk factor for respiratory disease (Jayes et al., 2016).

However, our investigations in both samples yielded some particularly unexpected findings: Compared to other clusters in the older sample, people in the *Abstainer but inactive* cluster had the highest prevalence of multimorbidity, complex multimorbidity, and endocrine disorders. Similarly, in the general adult sample, the *Overall low-risk* cluster had the highest prevalence of multimorbidity. These two clusters had similar sociodemographic profiles characterised by a two-thirds female majority, older age, and notable socioeconomic disadvantages (i.e. more likely to fall in lower levels of wealth/income and lower education levels), which are all well-established socio-demographic determinants of multimorbidity (Marmot et al., 2012). Yet, the persistence of these ostensibly counterintuitive relations between clusters of risk behaviours and disease outcomes despite accounting for key sociodemographic factors (i.e. age, sex, education, wealth, occupation) is what makes these findings particularly intriguing. Potential explanations for this observation are detailed in section 7.2.1. Nevertheless, the observation that, despite not engaging in risky behaviours, these clusters had a higher disease prevalence is consistent with a recent study on lifestyle risk factors for multimorbidity which found that among women, none of the unhealthy SNAP behaviours had an independent and statistically significant association with multimorbidity, but being older, less educated, and overweight were all associated with multimorbidity (de Almeida et al., 2020).

Within the general adult sample, I also examined morbid obesity as a separate health outcome and found that participants in the *Overall low-risk* cluster had a higher prevalence of morbid obesity compared to participants in other clusters. Obesity, typically measured by BMI, functions both as a predictor and an outcome of various health conditions. A high BMI (≥25)

is a well-established predictor of chronic diseases such as cardiovascular disease, diabetes, and certain cancers, primarily due to its role in metabolic disturbances and systemic inflammation (Delpino et al., 2023). However, obesity is also an outcome influenced by factors like diet, physical inactivity, socioeconomic status, and psychological stress (Fruh, 2017). This dual role complicates analysis, as obesity can both result from and contribute to the worsening of health conditions, creating a feedback loop that exacerbates multimorbidity (Delpino et al., 2023). Since obesity was only examined in the final waves of data collection (i.e. Waves 11,12,13), its contribution as a disease risk factor (i.e. predictor) could not be assessed in the UKHLS dataset. Nevertheless, the relatively high prevalence of morbid obesity in the *Overall low-risk* cluster, combined with socioeconomic disadvantage —both of which greatly increase the risk of eventually developing multimorbidity (Álvarez-Gálvez et al., 2023) — can partially explain the high multimorbidity prevalence in the *Overall low-risk* cluster.

The role of social determinants of health was further highlighted by considering the sociodemographic characteristics of participants in the clusters that had the lowest prevalence for most of the conditions studied. In both samples, these clusters had a distinctive profile: they were most likely to belong to the highest quintile of wealth, be educated to the highest level and least likely to belong to a minority ethnic group; and despite having low prevalence for the health conditions, they were all involved a high probability of engaging in one or more risky behaviour. These findings align with extensive evidence on the influence of socioeconomic status on the development and prevalence of chronic illness and multimorbidity, suggesting that individuals from more socioeconomically advantaged backgrounds are inherently less vulnerable to the adverse effects of unhealthy lifestyles than their more socioeconomically disadvantaged counterparts (Mair and Jani, 2020; Foster et al., 2018). Put another way, the present findings support a vulnerability hypothesis, whereby deprived groups are likely to

experience disproportionate harm from unhealthy lifestyles, which is not simply explained by a greater prevalence of unhealthy lifestyles (Foster et al., 2018). Although the mechanisms underpinning such disproportionate lifestyle-related harm in deprived groups remain unclear, some possibilities are explored in section 7.2.1.

7.1.4 Moderating role of sociodemographic characteristics on the relationship between behavioural clusters and disease

The research study presented in Chapter 6 built on the previous study of ELSA (Suhag, Webb and Holmes, 2024) to examine whether and how sociodemographic characteristics moderate the relationship between patterns of SNAP behaviours over time and disease outcomes, specifically multimorbidity, complex multimorbidity and disorders of six body systems. Rather than testing whether sociodemographic characteristics moderate the effect of single risk behaviours on disease outcomes (as traditional moderation analyses do; Brummett et al., 2003), the study used a multidimensional latent predictor, which represented subgroups of older adults with similar longitudinal patterns of SNAP behaviours. This approach allowed us to examine whether sociodemographic characteristics amplify or attenuate the association between multiple risk behaviours over time and chronic diseases — an examination that is often neglected as sociodemographic characteristics are typically controlled for in studies investigating the influence of risk behaviours on disease outcomes (e.g., Smyth et al., 2015).

Overall, there was little evidence to suggest that sociodemographic characteristics moderated the relationship between latent class membership and disease outcomes. The main effects of sociodemographic characteristics (i.e. age, sex, education, wealth and occupation) on disease

outcomes largely aligned with existing evidence (Singer et al., 2019a; Álvarez-Gálvez et al., 2023). For example, among the key socioeconomic indicators examined, the study found wealth but not education and occupation levels had a significant association with multimorbidity, complex multimorbidity, and endocrine disorders, after controlling for other measures of socioeconomic status. These findings are in line with a wealth of prior research indicating that material deprivation is associated with higher prevalence of multimorbidity and complex multimorbidity (Singer et al., 2019a; Álvarez-Gálvez et al., 2023). The study also found that, compared to males, females were more likely to have multimorbidity and complex multimorbidity. This finding also corroborates existing literature, including systematic reviews and meta-analyses, which consistently report a higher prevalence of multimorbidity among women (Nguyen et al., 2019; Agur et al., 2016; Violan et al., 2014).

Although the main effects aligned with prior research, the lack of significant moderation effects—likely due to statistical underpowering—prevents definitive conclusions. The wide confidence intervals in our moderation analysis suggest possible effects of a small sample size. Moreover, even in its most basic form a latent class moderation is an inherently complex model that requires careful interpretation. With the latent variable already simplifying the dynamics of multiple risk behaviours over time, incorporating multiple socio-demographic characteristics and health outcomes adds layers of complexity. This integration can pose a challenging for even a sufficiently well-powered statistical model in deciphering interactions and deriving sufficiently narrow confidence intervals. Given these challenges, the conclusions must be cautiously interpreted, recognising that the limited sample size and breadth of comparisons could have contributed to an underpowered analysis.

7.2 Implications and directions for future research

The findings of the studies presented in this thesis potentially have important implications for both research and policy, as discussed below.

7.2.1 Need to better understand of the dynamic between behavioural risk factors and social determinants of health

This thesis examined how SNAP behaviours clustered over time and their association with multimorbidity and a number of other disease outcomes. Although the studies allowed for a relatively short time between exposure (risk behaviours) and outcome (disease), it was observed that clusters characterised by individuals engaging in a higher number of risky behaviours were not associated with poor health outcomes to the degree we might have expected (Adams et al., 2017; Ng et al., 2020). Instead, the studies found that individuals in clusters with higher levels of socio-economic disadvantage (i.e. less wealthy and less well-educated) were more likely to experience worse health outcomes, despite engaging in fewer risky behaviours. This is broadly consistent with previous epidemiological studies which found that, although physical activity, diet, smoking, and socio-economic status independently affect the risk of multimorbidity (Agborsangaya et al., 2012; Aminisani et al., 2020; Mounce et al., 2018; Geda, Janzen and Pahwa, 2021; Luben et al., 2020; Nguyen et al., 2019), the relationship between these factors and subsequent accumulation of disease is complex and bidirectional (Suls, Green and Davidson, 2016).

Since the studies examined disease prevalence (the proportion of a population who have a specific characteristic in a given time period, regardless of when they first developed the

characteristic), not incidence (the number of new cases of a characteristic that develop in a population in a specified time period), the bi-directionality of the relationship between risk behaviours and disease outcomes could not be accounted for. Equally, although I examined risk behaviours across a decade-long period, the time-period between exposures (i.e. risk behaviours over time) and outcomes (i.e. disease status at the final wave) was relatively short and precludes causal attributions. Future studies that can examine incidence, with an adequate gap between measures of behaviour and disease outcomes, may be better able to account for the potential for bi-directionality and quantify disease risk attributable to risk behaviour patterns.

Unexpected differences in disease prevalence between clusters persisted after controlling for numerous, carefully selected socio-demographic covariates known to be associated with multimorbidity. There could be several explanations for this. First, despite controlling for several individual characteristics (i.e. age, sex, education, wealth, occupation), there may be underlying structural inequalities impacting outcomes that are hard to measure and record in longitudinal surveys (Liao, 2009). These might include disparities in access to resources, discrimination, historical inequities, and systemic biases within institutions (Lamont, Beljean and Clair, 2014; Liao, 2009). Given extensive research that shows that ethnic minorities bear the brunt of these structural inequalities and were overrepresented in the cluster with the least risky behaviour profile but worst health outcomes, the aforementioned structural inequalities could play a role. Second, the studies did not consider environmental and area-level factors that could influence health outcomes as our model was already quite complex and the longitudinal nature of our data meant our sample size was already truncated (Knies and Kumari, 2022). Thus, variables like the index of multiple deprivation were not included due to them having a significantly higher proportion of missing data (Ingram et al., 2021). Evidence

suggests that other factors like cultural norms, social networks, and community support systems can play a significant role in shaping outcomes and disparities in these outcomes even after accounting for demographic variables (Chuang, Chuang and Yang, 2013) and so may explain why differences in disease prevalence between clusters persisted after controlling for socio-demographic covariates. Fourth, sociodemographic categories intersect, meaning that individuals may face compounded disadvantages due to multiple aspects of their identity (e.g., race, gender, socioeconomic status). Therefore, controlling for one factor may not fully capture the complex interactions between different dimensions of identity and their effects on outcomes (Bauer, 2014). Finally, it is important to note that the average age of both samples was above 50 years, and despite accounting for key socio-economic variables at the baseline wave, epidemiological evidence shows that disease accumulation is not just a static process driven by biological depletion of body reserves but rather a dynamic state that is shaped by different social, psychological, and behavioural factors throughout the life course (Suls, Green and Davidson, 2016). For instance, some research shows that resilience to chronic disease may be built when an individual experiences certain socio-environmental stimuli during critical periods earlier in their life (Ben-Shlomo and Kuh, 2002). This includes evidence that low physical activity levels and financial hardship during childhood and adolescence, but not at other time points in the life course, can independently increase the risk of developing multimorbidity later in life (Feter et al., 2021; Tucker-Seeley et al., 2011).

The studies in this thesis uncovered some useful, albeit unexpected patterns, which seemed to highlight the role of socioeconomic disadvantage more than the number of risky behaviours engaged in. By examining whether and how clusters of behaviour changed over time and as individuals age, the research tried to adopt a life course approach. However, ultimately, the datasets—as well as the waves of data collection that included information on risk

behaviours— only covered specific periods of life, typically later life. Future research studies aiming to: i) further examine causal explanations for the different relative contributions of risk behaviours and sociodemographic factors to chronic disease development, or ii) identify optimal points of intervention or critical periods that influence disease outcomes, should therefore examine risk behaviours and social determinants across the life course (notably, including early in life) on associations between exposures and health outcomes at the individual and population levels (Lynch and Smith, 2005). Life course approaches focus on the particular way in which time and timing in relation to physical growth, reproduction, infection, social mobility, risk behaviour patterns, and transitions and so on, influence adult chronic diseases in different ways, and how these temporal processes interconnect and manifest in population-level disease trends (Ben-Shlomo and Kuh, 2002; Wagner et al., 2024). In practice, this might involve using samples spanning different life stages—from childhood to young adulthood—that can provide a nuanced understanding of when and how interventions should be targeted to mitigate disease burden effectively. It could also involve incorporating comprehensive measures of deprivation and attempting to use multi-level models to better explore the contribution of individual- versus area-level factors driving disparities in disease outcomes.

7.2.2 Emerging lifestyles, their longitudinal patterns and mental health outcomes

Given that epidemiology is the study of health and disease in populations in relation to their environment and ways of living (Morris, 1955), it follows that our research focus should adapt as these environments and lifestyles evolve. Therefore, I recommend that future studies investigating clusters of risk behaviours incorporate information on emerging risk factors such

as prolonged sitting, irregular sleep patterns, and excessive screen time (Wijndaele and Healy, 2016; Makarem et al., 2020; Liew and Aung, 2021; Neophytou, Manwell and Eikelboom, 2021). These behaviours not only affect physical health but also have significant implications for mental health (Yi et al., 2020; Rodriguez-Ayllon et al., 2019), which the present research could not examine as a disease outcome because of significant missing data on mental and behavioural disorders. Given the inherent link between physical and mental health (Ohrnberger, Fichera and Sutton, 2017), future studies might include measures of mental health as either a cause of poor physical health and / or an outcome of poor physical health.

Extending the point above about the need for studies to consider the full life course, it is also important to extend our understanding of how risk behaviours cluster over time and their association with adverse health outcomes in younger individuals (i.e. children, adolescents and young adults). Such studies can help to identify critical periods for intervention, which is vital given evidence that young people have more unstable risk patterns, tending towards less healthy profiles over time (Mahalik et al., 2013; Daw, Margolis and Wright, 2017) as well as scope for higher lifetime health impact if targeted early (Campbell et al., 2014). Examining data over decades could offer further insights into whether and how (clusters of) behaviour evolve and affect health outcomes across the life course, either in conjunction with or independent of social determinants.

7.2.3 Incorporating Biomarkers of Disease

Epidemiological research typically relies on a binary classification of disease, which is important for diagnostic purposes (i.e., people either do or do not have a disease and therefore do or do not qualify for particular treatment pathways). However, this overlooks the reality that

diseases often manifest along a continuum of severity rather than as distinct entities. For example, blood pressure varies from low to high, with intermediate levels also posing health risks (Brown et al., 2017). To prevent heart problems and strokes, it is recommended to address the risks across the entire range of blood pressure, not just when it is very high (Whelton, Bundy and Whelton, 2021). Similarly, epidemiological studies of cognitive function in elderly populations reveal that 'normality' imperceptibly merges into 'dementia', suggesting it is a quantitative, rather than qualitative or categorical phenomena and can pose risks across the entire range of presentations (Aisen et al., 2017). Thus, for the purposes of disease prevention, the whole continuum of risk is relevant and worth considering. Here, using biomarkers of disease may better gauge the impact of behavioural and social factors on multisystem biological risk, which may be difficult to assess by only examining cases versus non-cases. For example, inflammatory biomarkers have been linked to conditions like atherosclerosis and cardiovascular disease (Stoner et al., 2013). Higher resting metabolic rate per kilogram of lean body mass—a recognised biomarker of active resilience and accelerated ageing in animals and humans—has been found to predict future multimorbidity in older adults (Calderón-Larrañaga et al., 2019). Similarly, chronic systemic inflammation, assessed through biomarkers such as fibrinogen and C-reactive protein, has been implicated in various health problems and holds potential as a barometer of overall health status (Dhawan and Quyyumi, 2008). Incorporating this information could not only provide useful insights into the underlying physiological mechanisms associated with disease susceptibility but also aid in identifying at-risk populations for disease-prevention monitoring.

7.2.4 Longitudinal analysis of multiple risk behaviours

This thesis demonstrated a relative stability in risk behaviour patterns among UK adults, suggesting that longitudinal analyses of SNAP behaviours may offer limited additional insights beyond what cross-sectional studies can provide. However, this conclusion warrants careful consideration. The finding may not be universally applicable, particularly since the population studied primarily reflected patterns in older age groups, where lifestyle patterns tend to be more stable (Botoseneanu and Liang, 2012). In populations or settings where behaviours are more dynamic, such as among younger individuals, those experiencing rapid life transitions (e.g., students transitioning into the workforce), or people living in areas undergoing significant economic or social changes (e.g., communities affected by migration), we might observe more change in risk behaviours over time due to shifting social circumstances or access to resources, among other factors (Winpenny et al., 2018; Alidu and Grunfeld, 2018). These scenarios highlight the potential for more variable patterns of risk behaviours, where longitudinal analyses could offer valuable insights that cross-sectional studies might miss.

Furthermore, the landscape of public health is evolving, with emerging behaviours such as increased screen time and irregular sleep patterns potentially interacting with SNAP behaviours in complex ways (Yi et al., 2020; Rodriguez-Ayllon et al., 2019). These interactions may influence long-term health outcomes differently than what was observed in this thesis, indicating that the relevance of longitudinal studies may increase as these emerging risk behaviours become more prevalent. Thus, future research should expand longitudinal studies to encompass a broader range of populations and examine patterns of emerging health-risk behaviours. Additionally, cross-cultural studies could be instrumental in determining whether and which context-specific factors influence the stability of behaviours over time. In sum, while this thesis provides valuable insights into the population under study, further research is

essential to clarify when and where longitudinal analyses add significant value, ensuring that these findings are not overgeneralised to all populations or contexts.

7.3 Implications for policy

7.3.1 Prevention: high-risk versus population strategies

When the present research was conceptualised, one rationale for identifying clusters of health behaviour that were associated with elevated risk of disease was that these clusters could provide key information about high-risk groups that could be targeted in preventative interventions . However, the findings underscore the significant role of social disadvantages in the prevalence of disease. The discovery that the Overall low-risk clusters, characterised by the absence of risky behaviours, experienced worse outcomes underscores the need to consider risk behaviour patterns in the context of broader social determinants of health. As such, interventions targeting risk behaviours are likely to place a greater burden on individuals facing social disadvantages due to various barriers they encounter. In short, a narrow focus on behavioural clustering, without accounting for the complex interplay of social forces shaping individual behaviours, may inadvertently lead to counterproductive outcomes.

An alternative approach would be to situate findings from studies on risk behaviour clusters within the appropriate social context. That is, while measuring and monitoring behaviour can inform targeted health promotion strategies aimed at high-risk individuals, it could and perhaps should be complemented by a recognition of the structural influences on individual choices. In terms of interventions, the "high-risk preventative strategy [that] seeks to identify and help individuals who are either unduly susceptible or unusually exposed" does "not seek to alter the situations which determine exposure nor to attack the underlying reasons why the particular

health problem exists: it simply offers protection to the most vulnerable individuals from the hazardous problem, which continues" (Rose, 2008, p. 82). Our research underscores the persistent disparities in health outcomes, even after adjusting for sociodemographic factors, suggesting a potential influence of structural, cultural, and environmental factors. Thus, the high-risk preventative strategy ought to be coupled with population-wide strategies that aim to shift the risk distribution of the entire population in a more favourable direction — rather than just focusing on eliminating the high-risk tail—to yield greater impact (Rose, 2001). Recent attempts at modifying structural elements, such as the pricing, marketing, and availability of substances like alcohol and tobacco to affect population-level exposures, highlight the effectiveness of this balanced approach (Holmes et al., 2022; Horton et al., 2021). By framing interventions within this broader context, the root causes of health inequalities can be better addressed.

7.4 Critical reflections on the methodological approach of the thesis

Before discussing the strengths and limitations of the approach taken in the present thesis, I wanted to reflect on a few issues that influenced the choice of methods and approach. The key aim of the research presented in this thesis was to analyse clusters of multiple risk behaviours simultaneously over time to assess whether and how such clusters influence disease risk. This approach revealed unexpected findings to the extent that engaging in a higher number of risky behaviours was not always associated with worse health outcomes. Although the gap between exposures and outcomes was not sufficient to make any causal inferences, the present thesis nonetheless highlights vital connections that could be missed by solely focusing on health-risk

behaviour clusters and their socio-demographic correlates, without considering relevant health outcomes. Given that the majority of studies focusing on risk behaviour clusters to identify high-risk subgroups do not relate them to relevant outcomes, I recommend that including health outcomes could be key to deriving actionable insights from such analyses.

The use of repeated measures latent class analysis helped to simplify complexity by presenting information on various risk profiles, highlighting which socio-demographic subgroups may exhibit particular combinations of behaviour and associated disease outcomes. Yet, despite achieving some simplification, the findings revealed complex patterns that were challenging to distil into concise, straightforward messages. A potential solution could be to refine the scope of research questions and focus on specific subgroups, such as conducting the analysis on age-stratified samples. This approach could help unravel clearer and more actionable insights from the patterns observed. Nonetheless, grappling with such complexity is often an inherent aspect of conducting this type of analysis.

It is also crucial to approach specific behavioural clusters identified in individual studies with caution. Researchers' decisions on which behaviours to examine, how to categorise them, and which cluster analysis method to employ can all significantly influence results. Nevertheless, the replication of analysis and similarities in findings across different samples, varying in age, ethnic distribution, and measurement of behaviours, lends confidence to the identified patterns in the present thesis.

However, future researchers intending to explore multiple health risk behaviours would benefit from heeding the precautions that follow. First, measures of ostensibly the same variable can vary at different time points, even within the same longitudinal dataset. In my experience of

choosing which datasets to work with, I found that when a population study claimed to assess a risk behaviour over several waves, it was useful to first check if the measures used to examine that behaviour were consistent across waves before proceeding further. I faced this issue with a number of studies (e.g., Survey of Health, Ageing and Retirement in Europe (SHARE) dataset, first few waves of UKHLS) that had inconsistent cross-wave measures for some SNAP behaviours. Equally, although the initial plan was to examine risk behaviours across multiple decades, data on risk behaviours was sometimes not available for several waves or was assessed with measures that changed across some waves. A third aspect that took considerable time and effort involved decisions around which measures of behaviour to choose and how to categorise risky levels of behaviour (i.e. how many categories to divide each behaviour into and what cut-offs to use). I handled this by conducting an online survey to gather consensus from an opportunistic sample of experts, regarding the suitable choice of health behaviour measures and their respective categories. Finally, when choosing the best fitting latent class model, it was helpful to validate potential models by doing a split-half replication, because the best-fitting model suggested in the initial assessment could sometimes prove to be unstable in the replication.

Overall, I believe exploratory analyses through unsupervised learning methods (i.e. using unlabelled data to find patterns or structure within that data) like latent class modelling remain valuable for uncovering risk clusters and their association with chronic disease, especially in the context of exploring emerging lifestyle and environmental risk factors. Non-random clustering may now be more detectable due to electronic medical records, an ageing population with multiple health conditions, and advancements in epidemiological tools and data availability. By allowing the mapping of multiple risk factors and their association with diseases, latent modelling approaches could also help identify connections between disparate

conditions driven by unexplored clusters of risk factors (behavioural, sociodemographic, environmental or a combination of the three), with a potential to inform answers about disease aetiology.

7.4.1 Strengths

A key strength of the research presented in this thesis is that it was a first to explore the longitudinal clustering of multiple risk behaviours over time and their association with multimorbidity and a number of chronic diseases. The studies were conducted using advanced modelling techniques like repeated latent class analysis (RMLCA) and the findings were extended to a latent class moderation analysis using the identified clusters. RMLCA is a robust, model-based, probabilistic approach and the results demonstrated stable results in split-half replication. Moreover, the analysis is reproducible, as the diagnostic criteria and programming codes can be accessed.

The analyses were conducted using nationally-representative panel data, that were longitudinally weighted to ensure broad representativeness to older adults across England and the general adult population of the UK. The ongoing nature of ELSA and UKHLS also presents an opportunity to extend and potentially enhance existing research. By utilising upcoming waves of data collection, the brief timeframe between exposures and outcomes in the current study can be effectively addressed, thus allowing for a more thorough understanding of disease outcomes in the context of risk behaviour clusters. Notably, the results were adjusted for baseline disease and a range of socio-demographic variables that may confound the relationship between health behaviour and outcomes. Finally, the replication of the analyses and the

consistency observed in the clusters, along with their associated sociodemographic profiles and disease outcomes, bolster the credibility of the results.

7.4.2 Limitations

Despite these strengths, the studies included in this thesis have some limitations. First, increasing the number of latent classes, thereby improving model fit, may limit generalizability beyond the specific sample used. However, in the present case, the reproducibility of subgroups in the two datasets considered helped bolster the generalisability and validity of the identified classes (Sinha, Calfee and Delucchi, 2021).

Further, both ELSA and UKHLS relied on self-report data, which can be subject to recall limitations and social desirability bias. For example, people may be motivated to report that they consume, for example, less alcohol than they actually do, or may forget how much alcohol they consumed (Del Boca and Darkes, 2003). Having said this, longitudinal analyses may be less susceptible to misclassification bias due to repeated measures across survey waves (Haine, Dohoo and Dufour, 2018). Another constraint arose from the significant amount of missing data on mental health outcomes, which limited our ability to incorporate them either as separate outcomes or in measuring multimorbidity in either dataset.

Building on these methodological considerations, our decision to use a binary classification of <5/5+ for fruit and vegetable consumption was primarily informed by the consensus from the expert survey, which strongly supported this threshold. This approach also aligns with public health guidelines, such as the widely recommended intake of at least five servings of fruits and vegetables per day (Appleton et al., 2018). While adopting this binary classification facilitates

direct comparison with other studies that have examined similar dietary patterns and clusters, as many of these studies have used the same or similar cut-offs, alternative cut-offs such as <3, 3-5, and >5 servings per day would have provide a more nuanced understanding of consumption patterns. However, in the present case, using these could have complicated comparisons with existing research and public health recommendations. Another option could have been to treat fruit and vegetable consumption as a continuous variable, allowing for more precise measurements. However, given I used latent class analysis, it was crucial to use categorical variables, and the binary classification was the most appropriate choice to maintain clarity in public health messaging while ensuring robust comparisons with other cluster studies.

Additionally, the measure for dietary risk (i.e. daily fruit and vegetable intake) was not comprehensive (Ocean, Howley and Ensor, 2019). The assessment of diet in both ELSA and the UKHLS primarily relies on weekly fruit and vegetable intake, which, while suitable for screening purposes, may not fully capture disease-attributable risks of diet (Ocean, Howley and Ensor, 2019). This limitation arises because diet quality encompasses both nutrient density and energy intake, and a sufficient intake of nutrient-dense foods (e.g. fruits and vegetables), may contribute little to energy intake (Wirt and Collins, 2009). As a result, this measure fails to address potential inadequacies in dietary intake, such as inadequate protein intake. Protein-energy malnutrition, which is common among older adults living in high-resource countries, is associated with a higher susceptibility to infectious diseases, poor wound healing, and decreased functionality (del Carmen Alvarez-Nuncio and Ziegler, 2024; van der Pols-Vijlbrief et al., 2014). Similarly, dietary measures used in ELSA and UKHLS overlook information on energy-dense, nutrient-poor ultra-processed foods that account for more than 50% of the total dietary energy consumed in the UK, thus presenting an incomplete picture of the disease risks associated with dietary behaviour (Rauber et al., 2018).

A more comprehensive assessment of diet can be observed in the Survey of Health, Ageing and Retirement in Europe (SHARE) dataset (Alves and Perelman, 2022). For example, diet in the SHARE dataset is assessed by the frequency of consuming fruits, vegetables, as well as protein-rich foods such as dairy, legumes, eggs, meat, fish, or poultry (Alves and Perelman, 2022). These measures are commonly used to study adherence to the Mediterranean diet, which is associated with a reduced risk of a range of chronic diseases (Galbete et al., 2018), a lower risk of multimorbidity (Freisling et al., 2020; Kyprianidou et al., 2021), and cardiovascular disease (Rosato et al., 2019). Another noteworthy method is the 110-item food frequency questionnaire (FFQ) employed in the Lifelines Cohort Study conducted in the Netherlands (Vinke et al., 2018). Drawing on the 2015 Dutch Dietary Guidelines and relevant literature, the study evaluates nine food categories that are known to have beneficial health effects (i.e. vegetables, fruits, whole grains, legumes, nuts, fish, oils, dairy products, coffee, and tea) and three food categories with negative health effects (i.e. red and processed meats, butter and hard margarines, and sugar-sweetened beverages), Similarly, to assess the consumption of ultra-processed foods, examples of useful tools include a web-based self-administered 24-hour dietary record employed by the NutriNet-Santé prospective cohort study (Srour et al., 2019) and a semi-quantitative food frequency questionnaire with 114 food items, which is used to measure dietary habits in the ELSA-Brasil cohort study (Silva et al., 2018).

Finally, the possibility of a bidirectional relationship between risk behaviours and disease outcomes cannot be discounted, yet could not be fully explored due to the relatively short gap between exposures and outcomes. I faced this limitation because I chose data waves that had consistent health behaviour measures, which ended up restricting the timeframe available for analysis.

7.5 Conclusion

Overall, the present findings have several important implications and have made significant contributions to understanding SNAP risk behaviours and disease outcomes in the UK context. The research presented in this thesis is the first to examine and identify several different clusters represented by distinct patterns of SNAP risk behaviours, with relative stability of behaviours over time, not only in older adults but also in the general adult population in the UK. Secondly, this thesis has shown that there are clear relationships between these clusters of behaviour over time and multimorbidity and a number of other chronic health conditions. The present research is also clear that this relationship was not a simple linear relationship between the number of risk behaviours and adverse health outcomes, as has sometimes been examined in previous studies (Myint et al., 2011; Khan et al., 2023). Thirdly, a strong social gradient in the relationship between risk behaviour profiles and health outcomes was observed. The clusters characterised by sociodemographic profiles associated with significant social disadvantages were more likely to have worse health outcomes – sometimes despite engaging in fewer risky behaviours than some other clusters that had the highest levels of education and wealth but engaged in riskier behaviour profiles. Finally, across the studies using ELSA and UKHLS, broad similarities were found in the behaviour profiles and sociodemographic characteristics of clusters that had both the highest and the lowest prevalence of multimorbidity and most health conditions studied.

Overall, these findings emphasise the need to consider broader social determinants of health and advocate for examining risk behaviours within the relevant social context and with reference to relevant disease outcomes —rather than a narrow focus on behavioural clustering

alone. Future studies can effectively integrate and expand on the current findings by taking a life course approach to studying a diverse range of risk factors (e.g., by including emerging risk behaviours like prolonged sitting, excessive screen time, poor sleep habits) and extend the examination of disease outcomes to mental health outcomes as well as clinically viable biomarkers of disease to better capture progression from health to disease in the population.

References

Academy of Medical Sciences 2018. Multimorbidity: A priority for global health research (2018).

Adams, M. L., Grandpre, J., Katz, D. L. & Shenson, D. 2017. Linear association between number of modifiable risk factors and multiple chronic conditions: Results from the Behavioral Risk Factor Surveillance System. *Preventive medicine,* 105**,** 169-175.

Adena, M. & Myck, M. 2014. Poverty and transitions in health in later life. *Social science & medicine,* 116**,** 202-210.

Agborsangaya, C. B., Lau, D., Lahtinen, M., Cooke, T. & Johnson, J. A. 2012. Multimorbidity prevalence and patterns across socioeconomic determinants: a cross-sectional survey. *BMC public health,* 12**,** 1-8.

Agur, K., McLean, G., Hunt, K., Guthrie, B. & Mercer, S. W. 2016. How does sex influence multimorbidity? Secondary analysis of a large nationally representative dataset. *International journal of environmental research and public health,* 13**,** 391.

Aisen, P. S., Cummings, J., Jack, C. R., Morris, J. C., Sperling, R., Frölich, L., Jones, R. W., Dowsett, S. A., Matthews, B. R. & Raskin, J. 2017. On the path to 2025: understanding the Alzheimer's disease continuum. *Alzheimer's research & therapy,* 9**,** 1-10.

Alidu, L. & Grunfeld, E. A. 2018. A systematic review of acculturation, obesity and health behaviours among migrants to high-income countries. *Psychology & health,* 33**,** 724-745.

Álvarez-Gálvez, J., Ortega-Martín, E., Carretero-Bravo, J., Pérez-Muñoz, C., Suárez-Lledó, V. & Ramos-Fiol, B. 2023. Social determinants of multimorbidity patterns: A systematic review. *Frontiers in Public Health,* 11**,** 1081518.

Alves, R. & Perelman, J. 2022. European mature adults and elderly are moving closer to the Mediterranean diet—a longitudinal study, 2013–19. *European Journal of Public Health,* 32**,** 600-605.

Aminisani, N., Stephens, C., Allen, J., Alpass, F. & Shamshirgaran, S. M. 2020. Socio-demographic and lifestyle factors associated with multimorbidity in New Zealand. *Epidemiology and health,* 42.

Appleton, K., Krumplevska, K., Smith, E., Rooney, C., McKinley, M. & Woodside, J. 2018. Low fruit and vegetable consumption is associated with low knowledge of the details of the 5-a-day fruit and vegetable message in the UK: findings from two cross-sectional questionnaire studies. *Journal of human nutrition and dietetics,* 31**,** 121-130.

Artaud, F., Sabia, S., Dugravot, A., Kivimaki, M., Singh-Manoux, A. & Elbaz, A. 2016. Trajectories of unhealthy behaviors in midlife and risk of disability at older ages in the Whitehall II Cohort Study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences,* 71**,** 1500-1506.

Asparouhov, T. & Muthén, B. 2014. Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus web notes,* 21**,** 1-22.

Audrain-McGovern, J. & Benowitz, N. 2011. Cigarette smoking, nicotine, and body weight. *Clinical Pharmacology & Therapeutics,* 90**,** 164-168.

Banks, E., Joshy, G., Korda, R. J., Stavreski, B., Soga, K., Egger, S., Day, C., Clarke, N. E., Lewington, S. & Lopez, A. D. 2019. Tobacco smoking and risk of 36 cardiovascular disease subtypes: fatal and non-fatal outcomes in a large prospective Australian study. *BMC medicine,* 17**,** 1-18.

Banks, J., Phelps, A., Oskala, A., Steptoe, A., Blake, M., Oldfield, Z., Marmot, M., Clemens, S., Rogers, N., Nazroo, J. 2021. English Longitudinal Study of Ageing: Waves 0-9, 1998-2019.

Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S. & Guthrie, B. 2012. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet,* 380**,** 37-43.

Barta, J. A., Powell, C. A. & Wisnivesky, J. P. 2019. Global epidemiology of lung cancer. *Annals of global health,* 85.

Bauer, G. R. 2014. Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. *Social science & medicine,* 110**,** 10-17.

Bauer, U. E., Briss, P. A., Goodman, R. A. & Bowman, B. A. 2014. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *The Lancet,* 384**,** 45-52.

Bécares, L. 2021. Health and socio-economic inequalities by sexual orientation among older women in the United Kingdom: findings from the UK household longitudinal study. *Ageing & Society,* 41**,** 2416-2434.

Ben-Shlomo, Y. & Kuh, D. 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. Oxford University Press.

Benzies, K. M., Wångby, M. & Bergman, L. R. 2008. Stability and change in health-related behaviors of midlife Swedish women. *Health Care for Women International,* 29**,** 997-1018.

Birch, J., Petty, R., Hooper, L., Bauld, L., Rosenberg, G. & Vohra, J. 2019. Clustering of behavioural risk factors for health in UK adults in 2016: a cross-sectional survey. *Journal of Public Health,* 41**,** e226-e236.

Bolin, K. & Lindgren, B. 2016. Non-monotonic health behaviours–implications for individual health-related behaviour in a demand-for-health framework. *Journal of Health Economics,* 50**,** 9-26.

Bondy, S. & Rehm, J. 1998. The interplay of drinking patterns and other determinants of health. *Drug and Alcohol Review,* 17**,** 399-411.

Botoseneanu, A. & Liang, J. 2012. The effect of stability and change in health behaviors on trajectories of body mass index in older Americans: a 14-year longitudinal study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences,* 67**,** 1075-1084.

Braveman, P. & Gottlieb, L. 2014. The social determinants of health: it's time to consider the causes of the causes. *Public health reports,* 129**,** 19-31.

Bricca, A., Jäger, M., Johnston, M., Zangger, G., Harris, L. K., Midtgaard, J. & Skou, S. T. 2023. Effect of in-person delivered behavioural interventions in people with multimorbidity: systematic review and meta-analysis. *International Journal of Behavioral Medicine,* 30**,** 167-189.

Britton, A., Ben-Shlomo, Y., Benzeval, M., Kuh, D. & Bell, S. 2015. Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies. *BMC medicine,* 13**,** 1-9.

Brown, J. M., Robinson-Cohen, C., Luque-Fernandez, M. A., Allison, M. A., Baudrand, R., Ix, J. H., Kestenbaum, B., De Boer, I. H. & Vaidya, A. 2017. The spectrum of subclinical primary aldosteronism and incident hypertension: a cohort study. *Annals of internal medicine,* 167**,** 630-641.

Brummett, B. H., Babyak, M. A., Siegler, I. C., Mark, D. B., Williams, R. B. & Barefoot, J. C. 2003. Effect of smoking and sedentary behavior on the association between depressive

symptoms and mortality from coronary heart disease. *The American journal of cardiology,* 92**,** 529-532.
Buck, D. & Frosini, F. 2012. Clustering of unhealthy behaviours over time. *London: The Kings Fund***,** 1-24.
Burke, V., Zhao, Y., Lee, A., Hunter, E., Spargo, R., Gracey, M., Smith, R., Beilin, L. & Puddey, I. 2007. Health-related behaviours as predictors of mortality and morbidity in Australian Aborigines. *Preventive medicine,* 44**,** 135-142.
Burns, D. M. 2003. Epidemiology of smoking-induced cardiovascular disease. *Progress in cardiovascular diseases,* 46**,** 11-29.
Calderón-Larrañaga, A., Vetrano, D. L., Ferrucci, L., Mercer, S., Marengoni, A., Onder, G., Eriksdotter, M. & Fratiglioni, L. 2019. Multimorbidity and functional impairment–bidirectional interplay, synergistic effects and common pathways. *Journal of internal medicine,* 285**,** 255-271.
Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E. & Pan, Y. 2014. Early childhood investments substantially boost adult health. *Science,* 343**,** 1478-1485.
Canizares, M., Hogg-Johnson, S., Gignac, M. A., Glazier, R. H. & Badley, E. M. 2018. Increasing trajectories of multimorbidity over time: birth cohort differences and the role of changes in obesity and income. *The Journals of Gerontology: Series B,* 73**,** 1303-1314.
Cano, M. Á., de Dios, M. A., Correa-Fernández, V., Childress, S., Abrams, J. L. & Roncancio, A. M. 2017. Depressive symptom domains and alcohol use severity among Hispanic emerging adults: Examining moderating effects of gender. *Addictive behaviors,* 72**,** 72-78.
Chaoqun, D., Hua, C., Yeqin, Y., Yi, L., Yumei, S. & Hongyu, S. 2022. Patterns of risky health behaviors and associations with chronic diseases among young adult nursing students: a latent class analysis. *Journal of Nursing Research,* 30**,** e243.
Chen, Q., Luo, W., Palardy, G. J., Glaman, R. & McEnturff, A. 2017. The efficacy of common fit indices for enumerating classes in growth mixture models when nested data structure is ignored: A Monte Carlo study. *Sage Open,* 7**,** 2158244017700459.
Chen, Y. H., Karimi, M. & Rutten-van Mölken, M. P. 2020. The disease burden of multimorbidity and its interaction with educational level. *PloS one,* 15**,** e0243275.
Chuang, Y.-C., Chuang, K.-Y. & Yang, T.-H. 2013. Social cohesion matters in health. *International journal for equity in health,* 12**,** 1-12.
Chudasama, Y. V., Khunti, K. K., Zaccardi, F., Rowlands, A. V., Yates, T., Gillies, C. L., Davies, M. J. & Dhalwani, N. N. 2019. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. *BMC medicine,* 17**,** 1-13.
Cleland, C., Ferguson, S., Ellis, G. & Hunter, R. F. 2018. Validity of the International Physical Activity Questionnaire (IPAQ) for assessing moderate-to-vigorous physical activity and sedentary behaviour of older adults in the United Kingdom. *BMC medical research methodology,* 18**,** 1-12.
Collaborators, G. R. F. 2017. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet (London, England),* 390**,** 1345.
Collerton, J., Davies, K., Jagger, C., Kingston, A., Bond, J., Eccles, M. P., Robinson, L. A., Martin-Ruiz, C., von Zglinicki, T. & James, O. F. 2009. Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. *Bmj,* 339.
Conry, M. C., Morgan, K., Curry, P., McGee, H., Harrington, J., Ward, M. & Shelley, E. 2011. The clustering of health behaviours in Ireland and their relationship with mental health, self-rated health and quality of life. *BMC public health,* 11**,** 1-10.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A. & Sallis, J. F. 2003. International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise,* 35**,** 1381-1395.

Daw, J., Margolis, R. & Wright, L. 2017. Emerging adulthood, emergent health lifestyles: Sociodemographic determinants of trajectories of smoking, binge drinking, obesity, and sedentary behavior. *Journal of health and social behavior,* 58**,** 181-197.

de Almeida, M. G. N., Nascimento-Souza, M. A., Lima-Costa, M. F. & Peixoto, S. V. 2020. Lifestyle factors and multimorbidity among older adults (ELSI-Brazil). *European Journal of Ageing,* 17**,** 521-529.

De Bourdeaudhuij, I. & Van Oost, P. 1999. A cluster-analytical approach toward physical activity and other health related behaviors. *Medicine & Science in Sports & Exercise*.

De Vries, H., van't Riet, J., Spigt, M., Metsemakers, J., van den Akker, M., Vermunt, J. K. & Kremers, S. 2008. Clusters of lifestyle behaviors: results from the Dutch SMILE study. *Preventive medicine,* 46**,** 203-208.

Del Boca, F. K. & Darkes, J. 2003. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction,* 98**,** 1-12.

del Carmen Alvarez-Nuncio, M. & Ziegler, T. R. 2024. Micronutrient status and protein-energy malnutrition in free-living older adults: a current perspective. *Current opinion in gastroenterology,* 40**,** 99-105.

Delpino, F. M., dos Santos Rodrigues, A. P., Petarli, G. B., Machado, K. P., Flores, T. R., Batista, S. R. & Nunes, B. P. 2023. Overweight, obesity and risk of multimorbidity: A systematic review and meta-analysis of longitudinal studies. *Obesity Reviews,* 24**,** e13562.

Demakakos, P., Marmot, M. & Steptoe, A. 2012. Socioeconomic position and the incidence of type 2 diabetes: the ELSA study. *European journal of epidemiology,* 27**,** 367-378.

Desmond, C., Seeley, J., Groenewald, C., Ngwenya, N., Rich, K. & Barnett, T. 2019. Interpreting social determinants: Emergent properties and adolescent risk behaviour. *PLoS One,* 14**,** e0226241.

Dhalwani, N. N., O'Donovan, G., Zaccardi, F., Hamer, M., Yates, T., Davies, M. & Khunti, K. 2016. Long terms trends of multimorbidity and association with physical activity in older English population. *International Journal of Behavioral Nutrition and Physical Activity,* 13**,** 1-9.

Dhalwani, N. N., Zaccardi, F., O'Donovan, G., Carter, P., Hamer, M., Yates, T., Davies, M. & Khunti, K. 2017. Association between lifestyle factors and the incidence of multimorbidity in an older English population. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences,* 72**,** 528-534.

Dhawan, S. S. & Quyyumi, A. A. 2008. Rheumatoid arthritis and cardiovascular disease. *Current atherosclerosis reports,* 10**,** 128-133.

Dias, J. G. 2006. Latent class analysis and model selection. *From data and information analysis to knowledge engineering.* Springer.

Diez Roux, A. V. 2004. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiologic reviews,* 26**,** 104-111.

DiSantostefano, J. 2009. International classification of diseases 10th revision (ICD-10). *The Journal for Nurse Practitioners,* 5**,** 56-57.

Divo, M. J., Martinez, C. H. & Mannino, D. M. 2014. Ageing and the epidemiology of multimorbidity. *European Respiratory Journal,* 44**,** 1055-1068.

Dumith, S. C., Muniz, L. C., Tassitano, R. M., Hallal, P. C. & Menezes, A. M. 2012. Clustering of risk factors for chronic diseases among adolescents from Southern Brazil. *Preventive medicine,* 54**,** 393-396.

Elias, P. & McKnight, A. 2003. Earnings, unemployment and the NS-SEC. *A Researcher's Guide to the National Statistics Socio-economic Classification. London: Sage*, 151-172.

Farley, A. C., Hajek, P., Lycett, D. & Aveyard, P. 2012. Interventions for preventing weight gain after smoking cessation. *Cochrane database of systematic reviews*.

Feng, Z., van den Broek, T., Perra, O., Cramm, J. M. & Nieboer, A. P. 2022. Longitudinal health behaviour patterns among adults aged≥ 50 years in China and their associations with trajectories of depressive symptoms. *Aging & Mental Health*, 1-10.

Feter, N., Leite, J. S., Umpierre, D., Caputo, E. L. & Rombaldi, A. J. 2021. Multimorbidity and leisure-time physical activity over the life course: a population-based birth cohort study. *BMC public health,* 21, 1-11.

Filozof, C., Fernandez Pinilla, M. & Fernández-Cruz, A. 2004. Smoking cessation and weight gain. *Obesity reviews,* 5, 95-103.

Finckh, A., Gilbert, B., Hodkinson, B., Bae, S.-C., Thomas, R., Deane, K. D., Alpizar-Rodriguez, D. & Lauper, K. 2022. Global epidemiology of rheumatoid arthritis. *Nature Reviews Rheumatology,* 18, 591-602.

Fortin, M., Haggerty, J., Almirall, J., Bouhali, T., Sasseville, M. & Lemieux, M. 2014. Lifestyle factors and multimorbidity: a cross sectional study. *BMC public health,* 14, 1-8.

Fortin, M., Stewart, M., Poitras, M.-E., Almirall, J. & Maddocks, H. 2012. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *The Annals of Family Medicine,* 10, 142-151.

Foster, H. M., Celis-Morales, C. A., Nicholl, B. I., Petermann-Rocha, F., Pell, J. P., Gill, J. M., O'Donnell, C. A. & Mair, F. S. 2018. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. *The Lancet Public Health,* 3, e576-e585.

Fraser, S. D., Roderick, P. J., Aitken, G., Roth, M., Mindell, J. S., Moon, G. & O'Donoghue, D. 2014. Chronic kidney disease, albuminuria and socioeconomic status in the Health Surveys for England 2009 and 2010. *Journal of Public Health,* 36, 577-586.

Freedman, D. M., Sigurdson, A. J., Rajaraman, P., Doody, M. M., Linet, M. S. & Ron, E. 2006. The mortality risk of smoking and obesity combined. *American journal of preventive medicine,* 31, 355-362.

Freisling, H., Viallon, V., Lennon, H., Bagnardi, V., Ricci, C., Butterworth, A. S., Sweeting, M., Muller, D., Romieu, I. & Bazelle, P. 2020. Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. *BMC medicine,* 18, 1-11.

French, S., Rosenberg, M. & Knuiman, M. 2008. The clustering of health risk behaviours in a Western Australian adult population. *Health promotion journal of Australia,* 19, 203-209.

Fromme, K., Corbin, W. R. & Kruse, M. I. 2008. Behavioral risks during the transition from high school to college. *Developmental psychology,* 44, 1497.

Fruh, S. M. 2017. Obesity: Risk factors, complications, and strategies for sustainable long-term weight management. *Journal of the American association of nurse practitioners,* 29, S3-S14.

Fu, F. 2004. Profile of risk factors of coronary artery disease patients living in cities in China. *JESF,* 2, 36-46.

Fuchs, V. R. 2017. Social determinants of health: caveats and nuances. *Jama,* 317, 25-26.

Fumagalli, L., Knies, G. & Buck, N. 2017. Understanding Society, The UK Household Longitudinal Study, Harmonised British Household Panel Survey (BHPS) User Guide.

Gabriel, K. P., Sternfeld, B., Colvin, A., Stewart, A., Strotmeyer, E. S., Cauley, J. A., Dugan, S. & Karvonen-Gutierrez, C. 2017. Physical activity trajectories during midlife and subsequent risk of physical functioning decline in late mid-life: the study of women's health across the nation (Swan). *Preventive Medicine,* 105**,** 287-294.

Galbete, C., Schwingshackl, L., Schwedhelm, C., Boeing, H. & Schulze, M. B. 2018. Evaluating Mediterranean diet and risk of chronic disease in cohort studies: an umbrella review of meta-analyses. *European journal of epidemiology,* 33**,** 909-931.

Geda, N. R., Janzen, B. & Pahwa, P. 2021. Chronic disease multimorbidity among the Canadian population: prevalence and associated lifestyle factors. *Archives of Public Health,* 79**,** 60.

Global Burden of Disease Study Collaborators 2015. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013 diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet,* 386**,** 743-800.

Global Burden of Disease Study Collaborators 2017. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet (London, England),* 390**,** 1345.

Goodhind, A., Gilchrist, K. & Memon, A. 2014. Clustering of unhealthy behaviours over time: An analysis of health and wellbeing surveys in Southeast England, 2003 and 2012Anjum Memon. *European Journal of Public Health,* 24.

Graham, H., Hutchinson, J., Law, C., Platt, L. & Wardle, H. 2016. Multiple health behaviours among mothers and partners in England: clustering, social patterning and intra-couple concordance. *SSM-population health,* 2**,** 824-833.

Griffin, B., Sherman, K. A., Jones, M. & Bayl-Smith, P. 2014. The clustering of health behaviours in older Australians and its association with physical and psychological status, and sociodemographic indicators. *Annals of Behavioral Medicine,* 48**,** 205-214.

Gulliford, M. C. & Green, J. M. 2024. Is multimorbidity a useful concept for public health? *The Lancet Public Health,* 9**,** e210-e211.

Guo, B., Aveyard, P., Fielding, A. & Sutton, S. 2009. Using latent class and latent transition analysis to examine the transtheoretical model staging algorithm and sequential stage transition in adolescent smoking. *Substance use & misuse,* 44**,** 2028-2042.

Hackett, R. A., Moore, C., Steptoe, A. & Lassale, C. 2018. Health behaviour changes after type 2 diabetes diagnosis: Findings from the English Longitudinal Study of Ageing. *Scientific reports,* 8**,** 1-8.

Hailey, V., Bloomberg, M., Hamer, M. & Fisher, A. 2023. Association between neighbourhood cohesion and physical activity trajectories during the COVID-19 pandemic using data from Understanding Society: The UK Household Longitudinal Study & COVID-19 sub-study. *Preventive Medicine Reports,* 35**,** 102392.

Haine, D., Dohoo, I. & Dufour, S. 2018. Selection and misclassification biases in longitudinal studies. *Frontiers in veterinary science,* 5**,** 99.

Hamer, M., de Oliveira, C. & Demakakos, P. 2014. Non-exercise physical activity and survival: English longitudinal study of ageing. *American journal of preventive medicine,* 47**,** 452-460.

Harrison, C., Britt, H., Miller, G. & Henderson, J. 2014. Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. *BMJ open,* 4**,** e004694.

Harrison, C., Fortin, M., van den Akker, M., Mair, F., Calderon-Larranaga, A., Boland, F., Wallace, E., Jani, B. & Smith, S. 2021. Comorbidity versus multimorbidity: Why it matters. SAGE Publications Sage UK: London, England.

He, J. & Fan, X. 2018. Latent class analysis. *Encycl. Personal. Individ. Differ,* 1**,** 1-4.

He, L., Biddle, S. J., Lee, J. T., Duolikun, N., Zhang, L., Wang, Z. & Zhao, Y. 2021. The prevalence of multimorbidity and its association with physical activity and sleep duration in middle aged and elderly adults: a longitudinal analysis from China. *International Journal of Behavioral Nutrition and Physical Activity,* 18**,** 77.

Head, A., Birkett, M., Fleming, K., Kypridemos, C. & O'Flaherty, M. 2024. Socioeconomic inequalities in accumulation of multimorbidity in England from 2019 to 2049: a microsimulation projection study. *The Lancet Public Health,* 9**,** e231-e239.

Head, A., Fleming, K., Kypridemos, C., Pearson-Stuttard, J. & O'Flaherty, M. 2021a. Multimorbidity: the case for prevention. *J Epidemiol Community Health,* 75**,** 242-244.

Head, A., Fleming, K., Kypridemos, C., Schofield, P. & O'Flaherty, M. 2020. Dynamics of multimorbidity in England between 2004 and 2019: a descriptive epidemiology study. *European Journal of Public Health,* 30**,** ckaa165. 818.

Head, A., Fleming, K., Kypridemos, C., Schofield, P., Pearson-Stuttard, J. & O'Flaherty, M. 2021b. Inequalities in incident and prevalent multimorbidity in England, 2004–19: a population-based, descriptive study. *The Lancet Healthy Longevity,* 2**,** e489-e497.

Health, D. o. 2016. Alcohol Guidelines Review – Report from the Guidelines Development Group to the UK Chief Medical Officer. .

Hensel, D., Todd, K. & Engs, R. 2014. College student's health, drinking and smoking patterns: what has changed in 20 years? *College student journal,* 48**,** 378-385.

Hiscock, R., Bauld, L., Amos, A., Fidler, J. A. & Munafò, M. 2012. Socioeconomic status and smoking: a review. *Annals of the New York Academy of Sciences,* 1248**,** 107-123.

Ho, I. S.-S., Azcoaga-Lorenzo, A., Akbari, A., Davies, J., Hodgins, P., Khunti, K., Kadam, U., Lyons, R., McCowan, C. & Mercer, S. W. 2022a. Variation in the estimated prevalence of multimorbidity: systematic review and meta-analysis of 193 international studies. *BMJ open,* 12**,** e057017.

Ho, I. S.-S., Azcoaga-Lorenzo, A., Akbari, A., Davies, J., Khunti, K., Kadam, U. T., Lyons, R. A., McCowan, C., Mercer, S. W. & Nirantharakumar, K. 2022b. Measuring multimorbidity in research: Delphi consensus study. *BMJ medicine,* 1.

Hoekstra, T. 2013. Applied latent class models for epidemiology.

Holden, S., Barnett, A. H., Peters, J. R., Jenkins-Jones, S., Poole, C. D., Morgan, C. L. & Currie, C. J. 2013. The incidence of type 2 diabetes in the United Kingdom from 1991 to 2010. *Diabetes, Obesity and Metabolism,* 15**,** 844-852.

Holford, T. R. 1992. Analysing the temporal effects of age, period and cohort. *Statistical methods in medical research,* 1**,** 317-337.

Holmes, J., Angus, C., Boyd, J., Buykx, P., Brennan, A., Gardiner, K., Hernandez Alava, M., Hughes, J., Johnston, A. & Little, S. 2022. Evaluating the impact of Minimum Unit Pricing in Scotland on people who are drinking at harmful levels.

Horton, M., Perman-Howe, P. R., Angus, C., Bishop, J., Bogdanovica, I., Brennan, A., Britton, J., Brose, L. S., Brown, J. & Collin, J. 2021. The SPECTRUM consortium: a new UK Prevention Research Partnership consortium focussed on the commercial determinants of health, the prevention of non-communicable diseases, and the reduction of health inequalities. *Wellcome open research,* 6.

Hsu, H.-C., Luh, D.-L., Chang, W.-C. & Pan, L.-Y. 2013. Joint trajectories of multiple health-related behaviors among the elderly. *International journal of public health,* 58**,** 109-120.

Hultgren, B. A., Turrisi, R., Cleveland, M. J., Mallett, K. A., Reavy, R., Larimer, M. E., Geisner, I. M. & Hospital, M. M. 2019. Transitions in drinking behaviors across the college years: A latent transition analysis. *Addictive behaviors,* 92**,** 108-114.

Hurcombe, R., Bayley, M. & Goodman, A. 2010. Ethnicity and alcohol: A review of the UK literature.

Hutchesson, M. J., Duncan, M. J., Oftedal, S., Ashton, L. M., Oldmeadow, C., Kay-Lambkin, F. & Whatnall, M. C. 2021. Latent class analysis of multiple health risk behaviors among Australian university students and associations with psychological distress. *Nutrients,* 13**,** 425.

Ingram, E., Ledden, S., Beardon, S., Gomes, M., Hogarth, S., McDonald, H., Osborn, D. P. & Sheringham, J. 2021. Household and area-level social determinants of multimorbidity: a systematic review. *J Epidemiol Community Health,* 75**,** 232-241.

Institute for Social and Economic Research & Essex, U. o. 2023. Understanding Society: Waves 1-13, 2009-2022 and harmonised BHPS: Waves 1-18, 1991-2009. UK Data Service.

IPAQ Research Committee 2005. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms. *http://www. ipaq. ki. se/scoring. pdf*.

Irizar, P., Gage, S. H., Fallon, V. & Goodwin, L. 2022. A latent class analysis of health risk behaviours in the UK Police Service and their associations with mental health and job strain. *BMC psychiatry,* 22**,** 426.

Jayes, L., Haslam, P. L., Gratziou, C. G., Powell, P., Britton, J., Vardavas, C., Jimenez-Ruiz, C., Leonardi-Bee, J., Dautzenberg, B. & Lundbäck, B. 2016. SmokeHaz: systematic reviews and meta-analyses of the effects of smoking on respiratory health. *Chest,* 150**,** 164-179.

Johnston, M. C., Crilly, M., Black, C., Prescott, G. J. & Mercer, S. W. 2019. Defining and measuring multimorbidity: a systematic review of systematic reviews. *European journal of public health,* 29**,** 182-189.

Karamouzian, M., Cui, Z., Hayashi, K., Debeck, K., Reddon, H., Buxton, J. A. & Kerr, T. 2024. Longitudinal polysubstance use patterns and non-fatal overdose: a repeated measures latent class analysis. *International Journal of Drug Policy***,** 104301.

Katikireddi, S. V., Skivington, K., Leyland, A. H., Hunt, K. & Mercer, S. W. 2017. The contribution of risk factors to socioeconomic inequalities in multimorbidity across the lifecourse: a longitudinal analysis of the Twenty-07 cohort. *BMC medicine,* 15**,** 1-10.

Kershaw, K. N., Mezuk, B., Abdou, C. M., Rafferty, J. A. & Jackson, J. S. 2010. Socioeconomic position, health behaviors, and C-reactive protein: a moderated-mediation analysis. *Health Psychology,* 29**,** 307.

Khan, T. A., Field, D., Chen, V., Ahmad, S., Mejia, S. B., Kahleová, H., Rahelić, D., Salas-Salvadó, J., Leiter, L. A. & Uusitupa, M. 2023. Combination of multiple low-risk lifestyle behaviors and incident type 2 diabetes: a systematic review and dose-response meta-analysis of prospective cohort studies. *Diabetes Care,* 46**,** 643-656.

King, K., Meader, N., Wright, K., Graham, H., Power, C., Petticrew, M., White, M. & Sowden, A. J. 2015. Characteristics of interventions targeting multiple lifestyle risk behaviours in adult populations: a systematic scoping review. *PloS one,* 10**,** e0117015.

Kingston, A., Comas-Herrera, A. & Jagger, C. 2018. Forecasting the care needs of the older population in England over the next 20 years: estimates from the Population Ageing and Care Simulation (PACSim) modelling study. *The Lancet Public Health,* 3**,** e447-e455.

Kingston, A., Robinson, L., Booth, H., Knapp, M., Jagger, C. & Project, M. 2018. Projections of multi-morbidity in the older population in England to 2035: estimates from the

Population Ageing and Care Simulation (PACSim) model. *Age and ageing,* 47**,** 374-380.
Kino, S., Bernabé, E. & Sabbah, W. 2017. Socioeconomic inequality in clusters of health-related behaviours in Europe: latent class analysis of a cross-sectional European survey. *BMC Public Health,* 17**,** 1-8.
Kloppenburg, M. & Berenbaum, F. 2020. Osteoarthritis year in review 2019: epidemiology and therapy. *Osteoarthritis and cartilage,* 28**,** 242-248.
Knies, G. & Kumari, M. 2022. Multimorbidity is associated with the income, education, employment and health domains of area-level deprivation in adult residents in the UK. *Scientific Reports,* 12**,** 7280.
Kojima, G., Iliffe, S., Jivraj, S. & Walters, K. 2020. Fruit and Vegetable Consumption and Incident Prefrailty and Frailty in Community-Dwelling Older People: The English Longitudinal Study of Ageing. *Nutrients,* 12**,** 3882.
Kyprianidou, M., Panagiotakos, D., Faka, A., Kambanaros, M., Makris, K. C. & Christophi, C. A. 2021. Adherence to the Mediterranean diet in Cyprus and its relationship to multi-morbidity: An epidemiological study. *Public Health Nutrition,* 24**,** 4546-4555.
Lachman, M. E., Teshale, S. & Agrigoroaei, S. 2015. Midlife as a pivotal period in the life course: Balancing growth and decline at the crossroads of youth and old age. *International journal of behavioral development,* 39**,** 20-31.
Lamont, M., Beljean, S. & Clair, M. 2014. What is missing? Cultural processes and causal pathways to inequality. *Socio-Economic Review,* 12**,** 573-608.
Lanza, S. T. & Cooper, B. R. 2016. Latent class analysis for developmental research. *Child Development Perspectives,* 10**,** 59-64.
Lanza, S. T., Flaherty, B. P. & Collins, L. M. 2003. Latent class and latent transition analysis. *Handbook of psychology***,** 663-685.
Lanza, S. T., Tan, X. & Bray, B. C. 2013. Latent class analysis with distal outcomes: A flexible model-based approach. *Structural equation modeling: a multidisciplinary journal,* 20**,** 1-26.
Lawder, R., Harding, O., Stockton, D., Fischbacher, C., Brewster, D. H., Chalmers, J., Finlayson, A. & Conway, D. I. 2010. Is the Scottish population living dangerously? Prevalence of multiple risk factors: the Scottish Health Survey 2003. *BMC Public Health,* 10**,** 1-13.
Lear, S. A., Hu, W., Rangarajan, S., Gasevic, D., Leong, D., Iqbal, R., Casanova, A., Swaminathan, S., Anjana, R. M. & Kumar, R. 2017. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. *The Lancet,* 390**,** 2643-2654.
Lehnert, T. & König, H. 2012. Effects of multimorbidity on health care utilization and costs. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz,* 55**,** 685-692.
Lenz, B. 2001. The transition from adolescence to young adulthood: a theoretical perspective. *The Journal of School Nursing,* 17**,** 300-306.
Leventhal, A. M., Huh, J. & Dunton, G. F. 2014. Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis. *Perspectives in Public Health,* 134**,** 331-338.
Liao, J., Mawditt, C., Scholes, S., Lu, W., Umeda, M., Muniz Terrera, G., Hao, Y. & Mejía, S. 2019. Similarities and differences in health-related behavior clustering among older adults in eastern and western countries: A latent class analysis of global aging cohorts. *Geriatrics & gerontology international,* 19**,** 930-937.
Liao, T. F. 2009. Conceptualizing and measuring structural inequality. *Center for Research on Inequalities and the Life Course*.

Liew, S. C. & Aung, T. 2021. Sleep deprivation and its association with diseases-a review. *Sleep medicine,* 77**,** 192-204.

Lippke, S., Nigg, C. R. & Maddock, J. E. 2012. Health-promoting and health-risk behaviors: theory-driven analyses of multiple health behavior change in three international samples. *International journal of behavioral medicine,* 19**,** 1-13.

Lounassalo, I., Salin, K., Kankaanpää, A., Hirvensalo, M., Palomäki, S., Tolvanen, A., Yang, X. & Tammelin, T. H. 2019. Distinct trajectories of physical activity and related factors during the life course in the general population: a systematic review. *BMC public health,* 19**,** 271.

Lovatt, M., Eadie, D., Meier, P. S., Li, J., Bauld, L., Hastings, G. & Holmes, J. 2015. Lay epidemiology and the interpretation of low-risk drinking guidelines by adults in the United Kingdom. *Addiction,* 110**,** 1912-1919.

Luben, R., Hayat, S., Wareham, N., Pharoah, P. P. & Khaw, K.-T. 2020. Sociodemographic and lifestyle predictors of incident hospital admissions with multimorbidity in a general population, 1999–2019: the EPIC-Norfolk cohort. *Bmj Open,* 10**,** e042115.

Lucini, D., Solaro, N., Lesma, A., Gillet, V. B. & Pagani, M. 2011. Health promotion in the workplace: assessing stress and lifestyle with an intranet tool. *Journal of medical Internet research,* 13**,** e1798.

Luk, J. W. & Tsoh, J. Y. 2010. Moderation of gender on smoking and depression in Chinese Americans. *Addictive Behaviors,* 35**,** 1040-1043.

Lv, J., Liu, Q., Ren, Y., Gong, T., Wang, S. & Li, L. 2011. Socio-demographic association of multiple modifiable lifestyle risk factors and their clustering in a representative urban population of adults: a cross-sectional study in Hangzhou, China. *International Journal of Behavioral Nutrition and Physical Activity,* 8**,** 1-13.

Lynch, J. & Smith, G. D. 2005. A life course approach to chronic disease epidemiology. *Annu. Rev. Public Health,* 26**,** 1-35.

Mahalik, J. R., Levine Coley, R., McPherran Lombardi, C., Doyle Lynch, A., Markowitz, A. J. & Jaffee, S. R. 2013. Changes in health risk behaviors for males and females from early adolescence through early adulthood. *Health psychology,* 32**,** 685.

Mair, F. S. & Jani, B. D. 2020. Emerging trends and future research on the role of socioeconomic status in chronic illness and multimorbidity. *The Lancet Public Health,* 5**,** e128-e129.

Makarem, N., Zuraikat, F. M., Aggarwal, B., Jelic, S. & St-Onge, M.-P. 2020. Variability in sleep patterns: an emerging risk factor for hypertension. *Current hypertension reports,* 22**,** 1-10.

Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B. & Fratiglioni, L. 2011. Aging with multimorbidity: a systematic review of the literature. *Ageing research reviews,* 10**,** 430-439.

Marmot, M., Allen, J., Bell, R., Bloomer, E. & Goldblatt, P. 2012. WHO European review of social determinants of health and the health divide. *The Lancet,* 380**,** 1011-1029.

Marrero, J. A., Fontana, R. J., Fu, S., Conjeevaram, H. S., Su, G. L. & Lok, A. S. 2005. Alcohol, tobacco and obesity are synergistic risk factors for hepatocellular carcinoma. *Journal of hepatology,* 42**,** 218-224.

Martens, E., Gonnissen, H., Gatta-Cherifi, B., Janssens, P. & Westerterp-Plantenga, M. 2015. Maintenance of energy expenditure on high-protein vs. high-carbohydrate diets at a constant body weight may prevent a positive energy balance. *Clinical Nutrition,* 34**,** 968-975.

Mawditt, C., Sacker, A., Britton, A., Kelly, Y. & Cable, N. 2016. The clustering of health-related behaviours in a British population sample: testing for cohort differences. *Preventive medicine,* 88**,** 95-107.

McAloney, K., Graham, H., Law, C. & Platt, L. 2013. A scoping review of statistical approaches to the analysis of multiple health-related behaviours. *Preventive medicine,* 56**,** 365-371.
McEvoy, L. K., Kritz-Silverstein, D., Barrett-Connor, E., Bergstrom, J. & Laughlin, G. A. 2013. Changes in alcohol intake and their relationship with health status over a 24-year follow-up period in community-dwelling older adults. *Journal of the American Geriatrics Society,* 61**,** 1303-1308.
McFall, S., Nandi, A. & Platt, L. 2017. Understanding society: UK household longitudinal study: user guide to ethnicity and immigration research. *Institute for Social and Economic Research*.
McGeoch, L. J., Ross, S., Massa, M. S., Lewington, S. & Clarke, R. 2023. Cigarette smoking and risk of severe infectious respiratory diseases in UK adults: 12-year follow-up of UK biobank. *Journal of Public Health,* 45**,** e621-e629.
McGovern, P. & Nazroo, J. Y. 2015. Patterns and causes of health inequalities in later life: a Bourdieusian approach. *Sociology of Health & Illness,* 37**,** 143-160.
McKnight, A. & Elias, P. 2003. Earnings, unemployment and the NS-SEC.
McLernon, D. J., Powell, J. J., Jugdaohsingh, R. & Macdonald, H. M. 2012. Do lifestyle choices explain the effect of alcohol on bone mineral density in women around menopause? *The American journal of clinical nutrition,* 95**,** 1261-1269.
Meader, N., King, K., Moe-Byrne, T., Wright, K., Graham, H., Petticrew, M., Power, C., White, M. & Sowden, A. J. 2016. A systematic review on the clustering and co-occurrence of multiple risk behaviours. *BMC public health,* 16**,** 1-9.
Meier, P. S., Holmes, J., Brennan, A. & Angus, C. 2021. Alcohol policy and gender: a modelling study estimating gender-specific effects of alcohol pricing policies. *Addiction,* 116**,** 2372-2384.
Meier, P. S., Meng, Y., Holmes, J., Baumberg, B., Purshouse, R., Hill-McManus, D. & Brennan, A. 2013. Adjusting for unrecorded consumption in survey and per capita sales data: quantification of impact on gender-and age-specific alcohol-attributable fractions for oral and pharyngeal cancers in Great Britain. *Alcohol and Alcoholism,* 48**,** 241-249.
Mello, F. W., Melo, G., Pasetto, J. J., Silva, C. A. B., Warnakulasuriya, S. & Rivero, E. R. C. 2019. The synergistic effect of tobacco and alcohol consumption on oral squamous cell carcinoma: a systematic review and meta-analysis. *Clinical oral investigations,* 23**,** 2849-2859.
Mercer, S., Furler, J., Moffat, K., Fischbacher-Smith, D. & Sanci, L. 2016. *Multimorbidity: technical series on safer primary care*, World Health Organization.
Mercer, S. W., Smith, S. M., Wyke, S., O'Dowd, T. & Watt, G. C. 2009. Multimorbidity in primary care: developing the research agenda. Oxford University Press.
Miller, M. L. & Hurd, Y. L. 2017. Testing the gateway hypothesis. *Neuropsychopharmacology,* 42**,** 985-986.
Mishra, V. K., Srivastava, S. & Murthy, P. 2021. Population attributable risk for multimorbidity among adult women in India: Do smoking tobacco, chewing tobacco and consuming alcohol make a difference? *PloS one,* 16**,** e0259578.
Morris, J. N. 1955. Uses of epidemiology. *British Medical Journal,* 2**,** 395.
Mounce, L. T., Campbell, J. L., Henley, W. E., Arreal, M. C. T., Porter, I. & Valderas, J. M. 2018. Predicting incident multimorbidity. *The Annals of Family Medicine,* 16**,** 322-329.
Muthén, B. & Muthén, L. 2017a. Handbook of item response theory. Chapman and Hall/CRC Boca Raton, FL, USA:.
Muthén, B. & Muthén, L. 2017b. Mplus. *Handbook of item response theory.* Chapman and Hall/CRC.

Myint, P. K., Smith, R. D., Luben, R. N., Surtees, P. G., Wainwright, N. W., Wareham, N. J. & Khaw, K.-T. 2011. Lifestyle behaviours and quality-adjusted life years in middle and older age. *Age and ageing,* 40**,** 589-595.

Nash, A. 2017. National population projections: 2016-based statistical bulletin. *Office for National Statistics (ONS) October2015 Available at https://www. ons. gov. uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2015-10-29 [Google Scholar]*.

National Institute for Health Clinical Excellence 2022. *Alcohol-use disorders: diagnosis and clinical management of alcohol-related physical complications*, National Institute for Health and Clinical Excellence (NICE).

Neophytou, E., Manwell, L. A. & Eikelboom, R. 2021. Effects of excessive screen time on neurodevelopment, learning, memory, mental health, and neurodegeneration: A scoping review. *International Journal of Mental Health and Addiction,* 19**,** 724-744.

Ng, R., Sutradhar, R., Yao, Z., Wodchis, W. P. & Rosella, L. C. 2020. Smoking, drinking, diet and physical activity—modifiable lifestyle risk factors and their associations with age to first chronic disease. *International journal of epidemiology,* 49**,** 113-130.

Nguyen, H., Chua, K.-C., Dregan, A., Vitoratou, S., Bayes-Marin, I., Olaya, B. & Prina, A. M. 2020. Factors associated with multimorbidity patterns in older adults in England: findings from the English Longitudinal Study of Aging (ELSA). *Journal of aging and health,* 32**,** 1120-1132.

Nguyen, H., Manolova, G., Daskalopoulou, C., Vitoratou, S., Prince, M. & Prina, A. M. 2019. Prevalence of multimorbidity in community settings: A systematic review and meta-analysis of observational studies. *Journal of comorbidity,* 9**,** 2235042X19870934.

Nichter, M., Nichter, M., Carkoglu, A., Lloyd-Richardson, E. & Network, T. E. R. 2010. Smoking and drinking among college students:"It's a package deal". *Drug and alcohol dependence,* 106**,** 16-20.

Niebuur, J., Vonk, J. M., Du, Y., de Bock, G. H., Lunter, G., Krabbe, P. F., Alizadeh, B. Z., Snieder, H., Smidt, N. & Boezen, M. 2023. Lifestyle factors related to prevalent chronic disease multimorbidity: A population-based cross-sectional study. *Plos one,* 18**,** e0287263.

Nigg, C. R. & Long, C. R. 2012. A systematic review of single health behavior change interventions vs. multiple health behavior change interventions among older adults. *Translational behavioral medicine,* 2**,** 163-179.

Noble, N., Paul, C., Turon, H. & Oldmeadow, C. 2015. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. *Preventive medicine,* 81**,** 16-41.

Northwood, M., Ploeg, J., Markle-Reid, M. & Sherifali, D. 2018. Integrative review of the social determinants of health in older adults with multimorbidity. *Journal of Advanced Nursing,* 74**,** 45-60.

Nulu, S. 2017. Neglected chronic disease: The WHO framework on non-communicable diseases and implications for the global poor. *Global public health,* 12**,** 396-415.

Nylund-Gibson, K. & Choi, A. Y. 2018. Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science,* 4**,** 440.

Ocean, N., Howley, P. & Ensor, J. 2019. Lettuce be happy: A longitudinal UK study on the relationship between fruit and vegetable consumption and well-being. *Social Science & Medicine,* 222**,** 335-345.

Ohrnberger, J., Fichera, E. & Sutton, M. 2017. The relationship between physical and mental health: A mediation analysis. *Social science & medicine,* 195**,** 42-49.

Opazo Breton, M., Gillespie, D., Pryce, R., Bogdanovica, I., Angus, C., Hernandez Alava, M., Brennan, A. & Britton, J. 2022. Understanding long-term trends in smoking in England, 1972–2019: an age–period–cohort approach. *Addiction,* 117**,** 1392-1403.

Opazo Breton, M. & Gray, L. A. 2023. An age-period-cohort approach to studying long-term trends in obesity and overweight in England (1992–2019). *Obesity,* 31**,** 823-831.

Palladino, R., Tayu Lee, J., Ashworth, M., Triassi, M. & Millett, C. 2016. Associations between multimorbidity, healthcare utilisation and health status: evidence from 16 European countries. *Age and ageing,* 45**,** 431-435.

Pandey, R., Kumar, N., Paroha, S., Prasad, R., Yadav, M., Jain, S. & Yadav, H. 2013. Impact of obesity and diabetes on arthritis: An update. *Health,* 5**,** 143.

Pannu, G., Zaman, S., Bhala, N. & Zaman, R. 2009. Alcohol use in South Asians in the UK. British Medical Journal Publishing Group.

Pastuszak, A., Lisowski, K., Lewandowska, J. & Buśko, K. 2014. Level of physical activity of physical education students according to criteria of the IPAQ questionnaire and the recommendation of WHO experts. *Biomedical Human Kinetics,* 6.

Pathirana, T. I. & Jackson, C. A. 2018. Socioeconomic status and multimorbidity: a systematic review and meta-analysis. *Australian and New Zealand journal of public health,* 42**,** 186-194.

Peeters, M., Zondervan-Zwijnenburg, M., Vink, G. & Van de Schoot, R. 2015. How to handle missing data: A comparison of different approaches. *European Journal of Developmental Psychology,* 12**,** 377-394.

Perna, L., Mielck, A., Lacruz, M. E., Emeny, R. T., Holle, R., Breitfelder, A. & Ladwig, K. H. 2012. Socioeconomic position, resilience, and health behaviour among elderly people. *International journal of public health,* 57**,** 341-349.

Picco, L., Achilla, E., Abdin, E., Chong, S. A., Vaingankar, J. A., McCrone, P., Chua, H. C., Heng, D., Magadi, H. & Ng, L. L. 2016. Economic burden of multimorbidity among older adults: impact on healthcare and societal costs. *BMC health services research,* 16**,** 1-12.

Piette, J. D. & Kerr, E. A. 2006. The impact of comorbid chronic conditions on diabetes care. *Diabetes care,* 29**,** 725.

Poortinga, W. 2007. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Preventive medicine,* 44**,** 124-128.

Porcu, M. & Giambona, F. 2017. Introduction to latent class analysis with applications. *The Journal of Early Adolescence,* 37**,** 129-158.

Prochaska, J. J., Nigg, C. R., Spring, B., Velicer, W. F. & Prochaska, J. O. 2010. The benefits and challenges of multiple health behavior change in research and in practice. *Preventive medicine,* 50**,** 26-29.

Quintiliani, L., Allen, J., Marino, M., Kelly-Weeder, S. & Li, Y. 2010. Multiple health behavior clusters among female college students. *Patient education and counseling,* 79**,** 134-137.

Rauber, F., Louzada, M. L. d. C., Steele, E. M., Millett, C., Monteiro, C. A. & Levy, R. B. 2018. Ultra-processed food consumption and chronic non-communicable diseases-related dietary nutrient profile in the UK (2008–2014). *Nutrients,* 10**,** 587.

Rebholz, C., Rüegg, C. S., Michel, G., Ammann, R., Von der Weid, N., Kuehni, C. & Spycher, B. 2012. Clustering of health behaviours in adult survivors of childhood cancer and the general population. *British journal of cancer,* 107**,** 234-242.

Rhodes, R. E., Janssen, I., Bredin, S. S., Warburton, D. E. & Bauman, A. 2017. Physical activity: Health impact, prevalence, correlates and interventions. *Psychology & Health,* 32**,** 942-975.

Rodriguez-Ayllon, M., Estévez-López, F., Cadenas-Sanchez, C., Gracia-Marco, L., Lubans, D. R., Ortega, F. B. & Esteban-Cornejo, I. 2019. Physical activity, sedentary behaviour

and mental health in young people: a review of reviews. *Adolescent Health and Wellbeing: Current Strategies and Future Trends*, 35-73.
Rosato, V., Temple, N. J., La Vecchia, C., Castellan, G., Tavani, A. & Guercio, V. 2019. Mediterranean diet and cardiovascular disease: a systematic review and meta-analysis of observational studies. *European journal of nutrition,* 58, 173-191.
Rose, G. 2001. Sick individuals and sick populations. *International journal of epidemiology,* 30, 427-432.
Rose, G. 2008. Prevention for individuals and the 'high-risk'strategy. *Rose's Strategy of Preventive Medicine*, 65-86.
Safiri, S., Kolahi, A. A., Cross, M., Hill, C., Smith, E., Carson-Chahhoud, K., Mansournia, M. A., Almasi-Hashiani, A., Ashrafi-Asgarabad, A. & Kaufman, J. 2021. Prevalence, deaths, and disability-adjusted life years due to musculoskeletal disorders for 195 countries and territories 1990–2017. *Arthritis & rheumatology,* 73, 702-714.
Sakakibara, B. M., Obembe, A. O. & Eng, J. J. 2019. The prevalence of cardiometabolic multimorbidity and its association with physical activity, diet, and stress in Canada: evidence from a population-based cross-sectional study. *BMC Public Health,* 19, 1-9.
Sakib, M. N., Shooshtari, S., St. John, P. & Menec, V. 2019. The prevalence of multimorbidity and associations with lifestyle factors among middle-aged Canadians: an analysis of Canadian Longitudinal Study on Aging data. *BMC public health,* 19, 1-13.
Salisbury, C., Johnson, L., Purdy, S., Valderas, J. M. & Montgomery, A. A. 2011. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *British Journal of General Practice,* 61, e12-e21.
Salman, A. & Sellami, M. 2019. Do older adults with multimorbidity meet the recommended levels of physical activity? An analysis of Scottish Health Survey. *International Journal of Environmental Research and Public Health,* 16, 3748.
Schafer, J. L. 1999. Multiple imputation: a primer. *Statistical methods in medical research,* 8, 3-15.
Schmidt, H., Mah, C. L., Cook, B., Hoang, S., Taylor, E., Blacksher, E., Goldberg, D. S., Novick, L., Aspradaki, A. A. & Tzoutzas, I. 2016. Chronic disease prevention and health promotion. *Public health ethics: Cases spanning the globe*, 137-176.
Schneider, S., Huy, C., Schuessler, M., Diehl, K. & Schwarz, S. 2009. Optimising lifestyle interventions: identification of health behaviour patterns by cluster analysis in a German 50+ survey. *European Journal of Public Health,* 19, 271-277.
Schuit, A. J., van Loon, A. J. M., Tijhuis, M. & Ocké, M. C. 2002. Clustering of lifestyle risk factors in a general adult population. *Preventive medicine,* 35, 219-224.
Scotto Rosato, N. & Baer, J. C. 2012. Latent class analysis: A method for capturing heterogeneity. *Social Work Research,* 36, 61-69.
Seong, H., Lashley, H., Bowers, K., Holmes, S., Fortinsky, R. H., Zhu, S. & Corazzini, K. N. 2022. Resilience in relation to older adults with multimorbidity: A scoping review. *Geriatric Nursing,* 48, 85-93.
Shan, J., Yin, R. & Panuthai, S. 2024. Body mass index and multimorbidity risk: A systematic review and dose-response meta-analysis. *Archives of Gerontology and Geriatrics*, 105418.
Shang, X., Peng, W., Wu, J., He, M. & Zhang, L. 2020. Leading determinants for multimorbidity in middle-aged Australian men and women: A nine-year follow-up cohort study. *Preventive Medicine,* 141, 106260.
Shankar, A., McMunn, A. & Steptoe, A. 2010. Health-related behaviors in older adults: relationships with socioeconomic status. *American journal of preventive medicine,* 38, 39-46.

Silva, F. M., Giatti, L., de Figueiredo, R. C., Molina, M. d. C. B., de Oliveira Cardoso, L., Duncan, B. B. & Barreto, S. M. 2018. Consumption of ultra-processed food and obesity: cross sectional results from the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) cohort (2008–2010). *Public health nutrition,* 21**,** 2271-2279.

Singer, L., Green, M., Rowe, F., Ben-Shlomo, Y., Kulu, H. & Morrissey, K. 2019a. Trends in multimorbidity, complex multimorbidity and multiple functional limitations in the ageing population of England, 2002–2015. *Journal of comorbidity,* 9**,** 2235042X19872030.

Singer, L., Green, M., Rowe, F., Ben-Shlomo, Y. & Morrissey, K. 2019b. Social determinants of multimorbidity and multiple functional limitations among the ageing population of England, 2002–2015. *SSM-population health,* 8**,** 100413.

Sinha, P., Calfee, C. S. & Delucchi, K. L. 2021. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Critical care medicine,* 49**,** e63-e79.

Sisti, L. G., Dajko, M., Campanella, P., Shkurti, E., Ricciardi, W. & De Waure, C. 2018. The effect of multifactorial lifestyle interventions on cardiovascular risk factors: a systematic review and meta-analysis of trials conducted in the general population and high risk groups. *Preventive Medicine,* 109**,** 82-97.

Skivington, K., Katikireddi, S., Leyland, A., Hunt, K. & Mercer, S. 2015. Risk factors for multimorbidity: A multilevel analysis of a longitudinal cohort from Scotland: Kathryn Skivington. *The European Journal of Public Health,* 25**,** ckv167. 020.

Skou, S. T., Mair, F. S., Fortin, M., Guthrie, B., Nunes, B. P., Miranda, J. J., Boyd, C. M., Pati, S., Mtenga, S. & Smith, S. M. 2022. Multimorbidity. *Nature Reviews Disease Primers,* 8**,** 48.

Smith, S. M., Soubhi, H., Fortin, M., Hudon, C. & O'Dowd, T. 2012. Managing patients with multimorbidity: systematic review of interventions in primary care and community settings. *Bmj,* 345.

Smyth, A., Teo, K. K., Rangarajan, S., O'Donnell, M., Zhang, X., Rana, P., Leong, D. P., Dagenais, G., Seron, P. & Rosengren, A. 2015. Alcohol consumption and cardiovascular disease, cancer, injury, admission to hospital, and mortality: a prospective cohort study. *The Lancet,* 386**,** 1945-1954.

Soley-Bori, M., Ashworth, M., Bisquera, A., Dodhia, H., Lynch, R., Wang, Y. & Fox-Rushby, J. 2021. Impact of multimorbidity on healthcare costs and utilisation: a systematic review of the UK literature. *British Journal of General Practice,* 71**,** e39-e46.

Spring, B., Moller, A. C. & Coons, M. J. 2012. Multiple health behaviours: overview and implications. *Journal of public health,* 34**,** i3-i10.

Srour, B., Fezeu, L. K., Kesse-Guyot, E., Allès, B., Méjean, C., Andrianasolo, R. M., Chazelas, E., Deschasaux, M., Hercberg, S. & Galan, P. 2019. Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Santé). *bmj,* 365.

Stoner, L., Lucero, A. A., Palmer, B. R., Jones, L. M., Young, J. M. & Faulkner, J. 2013. Inflammatory biomarkers for predicting cardiovascular disease. *Clinical biochemistry,* 46**,** 1353-1371.

Suhag, A. 2023. alisha392/Health_Behaviour_Clusters: Version 1.0.0. Zenodo.

Suhag, A., Webb, T. L. & Holmes, J. 2024. Longitudinal clustering of health behaviours and their association with multimorbidity in older adults in England: A latent class analysis. *Plos one,* 19**,** e0297422.

Suls, J., Green, P. A. & Davidson, K. W. 2016. A biobehavioral framework to address the emerging challenge of multimorbidity. *Psychosomatic medicine,* 78**,** 281-289.

Tamakoshi, A., Tamakoshi, K., Lin, Y., Yagyu, K., Kikuchi, S. & Group, J. S. 2009. Healthy lifestyle and preventable death: findings from the Japan Collaborative Cohort (JACC) Study. *Preventive medicine,* 48**,** 486-492.

Tassitano, R. M., Tenório, M. C. M., Cabral, P. C. & Silva, G. A. P. d. 2014. Clustering of physical inactivity and low fruit and vegetables intake and associated factors in young adults. *Revista de Nutrição,* 27**,** 25-44.

Tegegne, T. K., Islam, S. M. S. & Maddison, R. 2022. Effects of lifestyle risk behaviour clustering on cardiovascular disease among UK adults: latent class analysis with distal outcomes. *Scientific Reports,* 12**,** 17349.

Tinner, L., Wright, C., Heron, J., Caldwell, D., Campbell, R. & Hickman, M. 2021. Is adolescent multiple risk behaviour associated with reduced socioeconomic status in young adulthood and do those with low socioeconomic backgrounds experience greater negative impact? Findings from two UK birth cohort studies. *BMC public health,* 21**,** 1-13.

Tobias, M., Jackson, G., Yeh, L. C. & Huang, K. 2007. Do healthy and unhealthy behaviours cluster in New Zealand? *Australian and New Zealand journal of public health,* 31**,** 155-163.

Tran, P. B., Kazibwe, J., Nikolaidis, G. F., Linnosmaa, I., Rijken, M. & van Olmen, J. 2022. Costs of multimorbidity: a systematic review and meta-analyses. *BMC medicine,* 20**,** 234.

Tsimpida, D., Kontopantelis, E., Ashcroft, D. M. & Panagioti, M. 2022. The dynamic relationship between hearing loss, quality of life, socioeconomic position and depression and the impact of hearing aids: answers from the English Longitudinal Study of Ageing (ELSA). *Social psychiatry and psychiatric epidemiology,* 57**,** 353-362.

Tucker, J. S., Orlando, M. & Ellickson, P. L. 2003. Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. *Health Psychology,* 22**,** 79.

Tucker-Seeley, R. D., Li, Y., Sorensen, G. & Subramanian, S. 2011. Lifecourse socioeconomic circumstances and multimorbidity among older adults. *BMC public health,* 11**,** 1-9.

van den Broek, T. 2021. Early-life circumstances, health behavior profiles, and later-life health in Great Britain. *Journal of Aging and Health,* 33**,** 317-330.

van der Pols-Vijlbrief, R., Wijnhoven, H. A., Schaap, L. A., Terwee, C. B. & Visser, M. 2014. Determinants of protein–energy malnutrition in community-dwelling older adults: a systematic review of observational studies. *Ageing research reviews,* 18**,** 112-131.

Van Gaal, L. F., Mertens, I. L. & De Block, C. E. 2006. Mechanisms linking obesity with cardiovascular disease. *Nature,* 444**,** 875-880.

van Nieuwenhuijzen, M., Junger, M., Velderman, M. K., Wiefferink, K. H., Paulussen, T. W., Hox, J. & Reijneveld, S. A. 2009. Clustering of health-compromising behavior and delinquency in adolescents and adults in the Dutch population. *Preventive medicine,* 48**,** 572-578.

Vinke, P. C., Corpeleijn, E., Dekker, L. H., Jacobs Jr, D. R., Navis, G. & Kromhout, D. 2018. Development of the food-based Lifelines Diet Score (LLDS) and its application in 129,369 Lifelines participants. *European journal of clinical nutrition,* 72**,** 1111-1119.

Violan, C., Foguet-Boreu, Q., Flores-Mateo, G., Salisbury, C., Blom, J., Freitag, M., Glynn, L., Muth, C. & Valderas, J. M. 2014. Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. *PloS one,* 9**,** e102149.

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C. & Vandenbroucke, J. P. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet,* 370**,** 1453-1457.

Wadd, S. & Papadopoulos, C. 2014. Drinking behaviour and alcohol-related harm amongst older adults: analysis of existing UK datasets. *BMC research notes,* 7**,** 1-9.
Wagner, C., Carmeli, C., Jackisch, J., Kivimäki, M., van der Linden, B. W., Cullati, S. & Chiolero, A. 2024. Life course epidemiology and public health. *The Lancet Public Health,* 9**,** e261-e269.
Whelton, S. P., Bundy, J. D. & Whelton, P. K. 2021. Coronary atherosclerosis across the continuum of blood pressure. *American Journal of Hypertension,* 34**,** 799-800.
Whitaker, V., Oldham, M., Angus, C., Fairbrother, H. & Holmes, J. 2020. The problem with good news: how should public health actors respond when alcohol consumption declines? *Addiction (Abingdon, England),* 115**,** 1987.
Whitaker, V., Oldham, M., Boyd, J., Fairbrother, H., Curtis, P., Meier, P. & Holmes, J. 2021. Clustering of health-related behaviours within children aged 11–16: a systematic review. *BMC public health,* 21**,** 1-12.
Wijndaele, K. & Healy, G. N. 2016. Sitting and chronic disease: where do we go from here? *Diabetologia,* 59**,** 688-691.
Wikström, K., Lindström, J., Harald, K., Peltonen, M. & Laatikainen, T. 2015. Clinical and lifestyle-related risk factors for incident multimorbidity: 10-year follow-up of Finnish population-based cohorts 1982–2012. *European journal of internal medicine,* 26**,** 211-216.
Willadsen, T. G., Bebe, A., Køster-Rasmussen, R., Jarbøl, D. E., Guassora, A. D., Waldorff, F. B., Reventlow, S. & Olivarius, N. d. F. 2016. The role of diseases, risk factors and symptoms in the definition of multimorbidity–a systematic review. *Scandinavian journal of primary health care,* 34**,** 112-121.
Wilson, K., Senay, I., Durantini, M., Sánchez, F., Hennessy, M., Spring, B. & Albarracín, D. 2015. When it comes to lifestyle recommendations, more is sometimes less: a meta-analysis of theoretical assumptions underlying the effectiveness of interventions promoting multiple behavior domain change. *Psychological bulletin,* 141**,** 474.
Winpenny, E. M., van Sluijs, E. M., White, M., Klepp, K.-I., Wold, B. & Lien, N. 2018. Changes in diet through adolescence and early adulthood: longitudinal trajectories and association with key life transitions. *International Journal of Behavioral Nutrition and Physical Activity,* 15**,** 1-9.
Wirt, A. & Collins, C. E. 2009. Diet quality–what is it and does it matter? *Public health nutrition,* 12**,** 2473-2492.
Wiscott, R., Kopera-Frye, K. & Begovic, A. 2002. Binge drinking in later life: Comparing young-old and old-old social drinkers. *Psychology of Addictive Behaviors,* 16**,** 252.
Wister, A., Li, L., Whitmore, C., Ferris, J., Klasa, K. & Linkov, I. 2022. Multimorbidity resilience and health behaviors among older adults: A longitudinal study using the Canadian Longitudinal Study on Aging. *Frontiers in Public Health,* 10**,** 896312.
Wister, A. V., Levasseur, M., Griffith, L. E. & Fyffe, I. 2015. Estimating multiple morbidity disease burden among older persons: a convergent construct validity study to discriminate among six chronic illness measures, CCHS 2008/09. *BMC geriatrics,* 15**,** 1-12.
Xie, H., Li, J., Zhu, X., Li, J., Yin, J., Ma, T., Luo, Y., He, L., Bai, Y. & Zhang, G. 2022. Association between healthy lifestyle and the occurrence of cardiometabolic multimorbidity in hypertensive patients: a prospective cohort study of UK Biobank. *Cardiovascular Diabetology,* 21**,** 199.
Yi, X., Liu, Z., Qiao, W., Xie, X., Yi, N., Dong, X. & Wang, B. 2020. Clustering effects of health risk behavior on mental health and physical activity in Chinese adolescents. *Health and quality of life outcomes,* 18**,** 1-10.

Yong, A. G. & Pearce, S. 2013. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology,* 9**,** 79-94.

Zantinge, E. M., van den Berg, M., Smit, H. A. & Picavet, H. S. J. 2014. Retirement and a healthy lifestyle: opportunity or pitfall? A narrative review of the literature. *The European Journal of Public Health,* 24**,** 433-439.

Zghebi, S. S., Steinke, D. T., Carr, M. J., Rutter, M. K., Emsley, R. A. & Ashcroft, D. M. 2017. Examining trends in type 2 diabetes incidence, prevalence and mortality in the UK between 2004 and 2014. *Diabetes, Obesity and Metabolism,* 19**,** 1537-1545.

Zhu, Y., Steele, F. & Moustaki, I. 2017. A general 3-step maximum likelihood approach to estimate the effects of multiple latent categorical variables on a distal outcome. *Structural Equation Modeling: A Multidisciplinary Journal,* 24**,** 643-656.

Zwolinsky, S., Raine, G. & Robertson, S. 2016. Prevalence, co-occurrence and clustering of lifestyle risk factors among UK men. *Journal of Men's Health,* 12.