University of Milan
Department of Clinical Sciences and Community Health
PhD in Public Health Sciences
XXXV cycle – Academic Year 2021-2022

HEALTHY LIFESTYLE AND RISK OF MULTIMORBIDITY

PhD student: Giulia PEVERI – R12586
Supervisor: Professor Rino BELLOCCO
Co-supervisor: Professor Federico AMBROGI

ABSTRACT:

Background Life expectancy has increased worldwide going from 45.7 years in 1950 to 72.6 years in 2019. A subgroup of these, chronic diseases (i.e., health problems requiring ongoing management over a period of years or decades), may lead to challenges in patient care when they present concomitantly (i.e., as multimorbidity). Because clinical trials often exclude patients with multimorbidity and most guidelines do not provide recommendations for multimorbid patients, these challenges persist. Moreover, multimorbidity negatively affects quality of life and functional ability and accelerates mortality. Many studies have been published on the role of modifiable lifestyle factors on multimorbidity, i.e., of tobacco and alcohol consumption, being overweight or obese, having a poor diet, and a low physical activity level. To our knowledge, only one study combined them in a comprehensive total lifestyle score. Therefore, a gap remains regarding the role of multiple lifestyle habits combined on multimorbidity.

Aim Our aim was: i) to determine the patterns of multimorbidity of selected groups of diseases or conditions, chosen among the major causes of death. and ii) to estimate the effect of five important modifiable lifestyle behaviors on the morbidity and multimorbidity of the selected diseases or conditions.

Methods To define multimorbidity we considered all chronic causes of death among the 369 diseases, injuries, and impairments recorded in the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) publicly available databases. We ranked the causes of death by decreasing yearly rates and grouped them as follows: 1) Cardiovascular diseases, i.e., ischemic heart disease, stroke, and hypertensive heart disease; 2) Gastrointestinal tract (GIT) cancers (i.e., colorectal, gastric, liver, pancreatic, and esophageal cancers) and respiratory tract (RT) cancers (i.e., trachea, bronchus, and lung cancers); 3) Alzheimer disease and other dementias; 4) Chronic obstructive pulmonary disease. These four macro-groups of chronic diseases are together responsible for an average cause-specific crude mortality rate of 827 deaths per thousand inhabitants. Multimorbidity was defined as the occurrence of diseases from two different groups.

We used data from the Swedish National March Cohort (SNMC) to study the association of five lifestyle factors with multimorbidity. We developed a partial healthy lifestyle index (HLI) score for each of the lifestyle exposures and a total HLI score ranging from 0 (worst habits) to 20 (best habits). Four states (i.e., baseline, morbidity, multimorbidity, and death) were used to define a multi-state framework, and each transition was modelled individually with a parametric multi-state model. We estimated transition probability between states and hazard ratios (HRs) and 95% confidence intervals (CIs) for the exposures of interest.

Results During an average follow-up time of 18.2 years we observed 6,458 morbidity cases, 946 transitions to multimorbidity, and 4,441 deaths. For values of the five partial HLI scores corresponding to healthier lifestyles we observed a reduction in the risk of morbidity, multimorbidity, and mortality, and we found that, e.g., over 15 years of follow-up, a man aged 65 years at baseline with an excellent lifestyle (all partial scores 4) would have a

33% reduction in the cumulative probability of morbidity, multimorbidity, and death combined compared to another man with same characteristics but a poor lifestyle (all partial scores 0-1). In case of women, the cumulative probability would be reduced by 29%. One unit increase in the total HLI score corresponded to 4% reduction in the risk of morbidity (HR [95% CI]: 0.96 [0.95-0.97]) and 6% in the risk of multimorbidity (0.94 [0.92-0.96]), similarly for both males and females. Having an HLI score of 16-20 halved the risk of morbidity compared to an HLI of 0-4 (HR [95% CI]: 0.47 [0.36-0.61] in men, and 0.46 [0.33-0.64] in women) with a stronger effect for women (p-value for heterogeneity 0.01), and reduced the risk of multimorbidity by two thirds (0.35 [0.20-0.63] in men and 0.30 [0.16-0.56] in women; p-value for heterogeneity <0.01).

Conclusions We found that healthy lifestyle habits, summarized by the HLI score, were inversely associated with morbidity and multimorbidity of selected cardiovascular diseases, gastrointestinal and respiratory cancers, dementia, and COPD. We determined that being a never smoker or quitting smoking, having a low alcohol consumption, high physical activity levels, and a low BMI, and following the Mediterranean Diet recommendations can lower the probability of morbidity, multimorbidity and death. This effect is particularly evident when all the healthy lifestyles are combined.

LIST OF ABBREVIATIONS:

AIC: Akaike's Information Criterion

AIDS: Acquired Immune Deficiency Syndrome

BMI: Body Mass Index

CI: Confidence Interval

CIF: Cumulative Incidence Function

COPD: Chronic Obstructive Pulmonary Disease

EPIC: European Prospective Investigation into Cancer and Nutrition

FFQ: Food Frequency Questionnaire

GIT: gastrointestinal tract

HHD: Hypertensive Heart Disease

HIV: Human Immunodeficiency Virus

HLI: Healthy Lifestyle Index

HR: Hazard Ratio

ICD: International Classification of Diseases

IHD: Ischemic Heart Disease

IPR: National Inpatient Register

IQR: Interquartile Range

MDP: Mediterranean Dietary Pattern

MET: Metabolic Equivalent of Task

MSM: Multi-state Model

SNMC: Swedish National March Cohort

NPR: National Patient Register

PIN: Personal Identity Number

PRTax: Population Register maintained by the Swedish Tax Agency

RT: respiratory tract

SNAC-K: Swedish National Study on Aging and Care in Kungsholmen

SD: Standard Deviation

TPR: Total Population Register

WHO: World Health Organization

Index

1 Background

1.1 Chronic diseases

Life expectancy has steeply increased worldwide in the last decades, going from just 45.7 years in 1950 to 72.6 years in 2019 (United Nations, Department of Economic and Social Affairs, Population Division 2020). In the same year, in Europe there were almost 103 million (20%) people aged 65 or more, and those aged 80 or more were more than 29 million (5.7%) (European Commission 2019). At the same time, communicable diseases have been replaced by non-communicable diseases as the predominant burden in health care: in 2016, non-communicable diseases caused an estimated 40.5 million (71%) of the 56.9 million deaths worldwide, the risk of dying from a non-communicable disease being highest in low- and middle-income countries (Bennett et al. 2018).

Chronic diseases are defined by the WHO as health problems requiring ongoing management over a period of years or decades (World Health Organization 2002), and comprise mainly non-communicable diseases. Cardiovascular diseases, cancer, diabetes, and chronic respiratory diseases are widely considered the main chronic diseases; additionally, since survival rates have improved and illness duration increased, certain mental disorders and disabilities, and even HIV infection have also progressively been included in the pool of chronic conditions (Busse et al. 2010; Deeks, Lewin, and Havlir 2013).

Chronic diseases, when combined, create additional challenges to patient care, since clinical trials usually exclude patients with coexisting conditions, and therefore most guidelines do not provide recommendations for patients presenting with multiple diseases (Lugtenberg et al. 2011). Moreover, polypharmacy may lead to adverse events and poor adherence to therapy (Gnjidic et al. 2012).

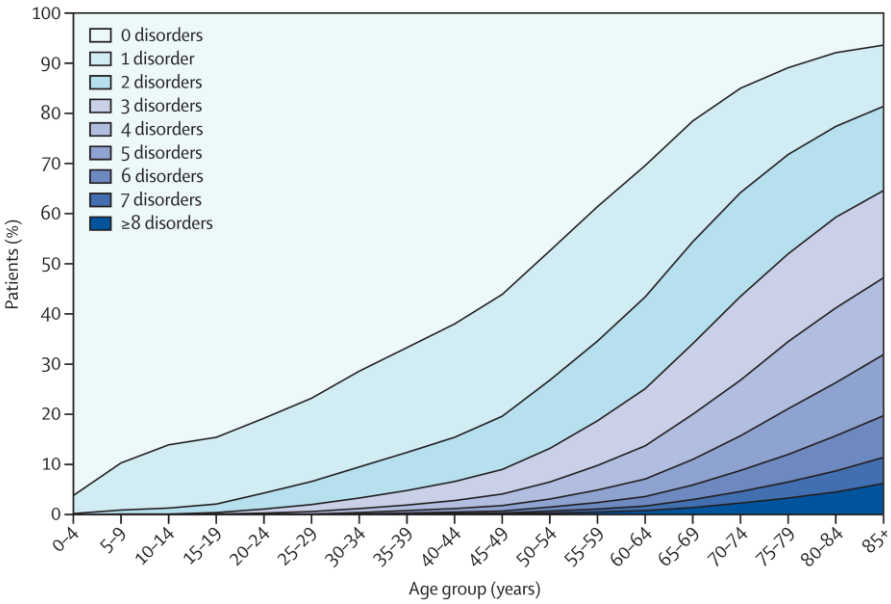
1.2 Multimorbidity

The prevalence of multiple chronic conditions appears to be heterogeneous among studies, depending on which diseases are included in the definition (Marengoni et al. 2011). Nevertheless, a Scottish study based on 1,751,841 patients (about a third of the Scottish population) found that 23.2% of the included subjects had at least two chronic disorders from a pool of 40 (Barnett et al. 2012). As shown in **Figure 1.1**, the number of multiple conditions increased with age.

To address the coexistence of two or more diseases or conditions, two terms are most commonly used. The first is comorbidity, coined in 1970 and defined as the presence of other diseases beyond an index disease (Feinstein 1970). It was introduced to evaluate the possible impact that additional conditions may have on the prognosis and therapeutic pathway of the disease of main interest. In fact, it was also used with the general meaning of "concomitant diseases" until the concept of multimorbidity, a term which also was already sporadically used, was

brought to the attention of the international scientific community to specifically define any co-occurrence of diseases, without prioritizing one disease in particular (van den Akker, Buntinx, and Knottnerus 1996).

Figure 1.1 Reproduction of the figure *"Number of chronic disorders by age-group"* (Barnett et al. 2012)



A systematic review on the use and definition of terms to describe the presence of multiple concurrent diseases identified a total of 461 papers including the term multimorbidity, but also 144 in which comorbidity was used without referring to an index disease, and then a variety of other terms: polypathology was found in 31 publications, polymorbidity in 31, multipathology in 6, multicondition in 5, and pluripathology in 3 (Almirall and Fortin 2013). About half of those using the term multimorbidity did not include any definition of the word; in the remaining half, thirteen essentially different definitions were given, the most frequent implying more than one chronic or long-term diseases/conditions and the second most frequent implying more than one disease or conditions, with no other specification.

Given that the traditional definition of multimorbidity proved to be inconsistent between researchers, some attempts have been made to identify critical issues in defining multimorbidity and to suggest a novel comprehensive definition.

The authors of a systematic review published in 2016 found that, of the 163 articles included in their research, 115 (71%) used individually constructed definitions of multimorbidity, and that all of these explicitly named in their definitions one or more diseases (e.g., diabetes, stroke, cancer), possibly in conjunction with risk factors (e.g., hypertension, osteoporosis, obesity) or symptoms (e.g., back pain, visual impairment, alcohol disorder) (Willadsen et al. 2016). Duration and severity of the diseases, risk factors or symptoms were also used as criteria in the definition of respectively 28% and 23% of the articles. Finally, the cutoff of two or more diseases to define multimorbidity was applied to different pools of possible conditions to choose from, ranging from 4 to 147.

The purpose of giving a comprehensive definition of multimorbidity was pursued by the authors of a systematic review published in 2013 (Le Reste et al. 2013). Again, the authors found a wide variety of definitions for multimorbidity, ranging from very simple (i.e., "comorbidity") to very complex (i.e., "overall impact of the different diseases in an individual, taking into account their severity and other health-related attributes or non–health-related individual attributes"). To achieve the aim of the study, the authors grouped the different criteria used in the literature to define multimorbidity into 11 thematic groups, which were: chronic disease, acute disease, biopsychosocial factors and somatic risk factors, coping strategies of the patient, burden of disease, health care consumption, disability, quality of life, frailty, social network, and health outcome.

The ultimate three-sentence definition proposed by Le Reste and colleagues, which would embrace all the definitions reviewed, is the following:

> "Multimorbidity is defined as any combination of chronic disease with at least one other disease (acute or chronic) or biopsychosocial factor (associated or not) or somatic risk factor.
>
> Any biopsychosocial factor, any somatic risk factor, the social network, the burden of diseases, the health care consumption, and the patient's coping strategies may function as modifiers (of the effects of multimorbidity).
>
> Multimorbidity may modify the health outcomes and lead to an increased disability or a decreased quality of life or frailty."

As illustrated by the authors, the first sentence reveals the actual definition of multimorbidity, the second is indicative of possible modifiers of the burden of multimorbidity for long-term health professionals and patients, and the third clarifies the outcomes of multimorbidity.

Table 1.1 shows an adaptation of the themes and subthemes identified by Le Reste and colleagues, used to create this definition.

This work was deemed useful by the authors of an umbrella review and meta-analysis conducted in 2017 on the definitions and measures of multimorbidity (Johnston et al. 2019). The authors appreciated the holistic perspective of the definition given by Le Reste and colleagues. However, they stated that a simple disease count, or measures (e.g., the Charlson index, the Index of Coexistent Disease, the Cumulative Illness Rating Scale) initially developed as comorbidity measures but increasingly used as multimorbidity measures, are particularly appropriate for certain outcomes. They suggested to consider including validated weighted measures when predicting the influence of multimorbidity on single outcomes, as they can be more appropriate or informative, and to choose disease count in case of multiple outcomes or populations, or where the evidence for a different measure is weak.

Table 1.1 Adaptation of the table *"Themes and Subthemes Identified for Multimorbidity Conditions"* (Le Reste et al. 2013)

Themes	Subthemes
Chronic disease	Chronic condition – Chronic diseases – Psychosomatic diseases/physical implications – Complexity characteristics of chronic disease
Acute disease	Acute conditions – Acute disease – Reaction to severe stress and adjustment disorders – Complexity characteristics of acute disease
Biopsychosocial factors and somatic risk factors	Somatic risk factors – Psychological risk factors – Psychosocial risk factors – Lifestyle – Demography: age, sex – Psychological distress – Sociodemographic characteristics – Aging – Patients beliefs/expectations – Physiology – Physiopathology
Coping	Patients' coping strategies
Burden of diseases	Disease complication – Disease morbidity
Health care consumption	Use of carers – Treatment or medication – Management – Disease management – Medical procedure – Malpractice – Health care services – Health care – Health care policy – Medical history – Family history – Assessment – Prevention – Pain – Health services/setting/treatment – Symptoms/signs/complaints – Cost of care – Polypharmacy
Disability	Handicap – Functional impairments
Quality of life	Quality of life – Health status – Impairment – Morbidity implications
Frailty	Frailty
Social network	Social network
Health outcomes	Mortality – Indicator – Outcome – Medical research epidemiology/instruments/level of multimorbidity – Classification of morbidity statistics

Moreover, Le Reste and colleagues, while recommending an ultimate definition of multimorbidity that would solve the problem of its heterogeneity in the literature, fail to give also an operative set of rules that researchers could use in order to apply this definition. Multimorbidity remains tied to the setting of the research, the clinical questions, and ultimately to the data available for the analysis.

1.2.1 Beyond multimorbidity: the concept of syndemic

In latest years, syndemic, a new concept close to that of multimorbidity, has been discussed with increased frequency and interest. Syndemic identifies clusters of social and health problems at a population level and its definition is based on: i) the presence of a cluster of at least two diseases or health conditions within a specific population; ii) the presence of social factors involved in the clustering; and iii) the clustering of diseases affects diseases interaction, leading to a higher burden of affected populations in terms of overall health (Singer et al. 2017). The first example of syndemic involves substance abuse, violence, and AIDS, and the term was coined to extend the concept of epidemic to factors other than those strictly related to a population's health (Singer 1996). Syndemics go beyond describing the co-occurrence of diseases and aim at grasping how and why the changes in the severity or progression of diseases happen.

1.3 Risk factors

Among all risk factors possibly involved in the development of multimorbidity, those related to lifestyle such as tobacco and alcohol consumption, being overweight or obese, having a poor diet, and being physically inactive are responsible for many years lost due to disability and a great number of premature death (Gakidou et al. 2017). Based on the recommendation of the WHO, a BMI in the range 18.5-24.9 kg/m^2 is considered optimal, together with at least 30 minutes of physical activity on most days of the week (Report of the Joint WHO/FAO expert consultation 2003). A more recent report by the WHO changed these recommendations to at least 150-300 minutes of moderate-intensity aerobic physical activity or at least 75-150 minutes of vigorous-intensity aerobic physical activity per week, plus muscle-strengthening activities at least twice a week (*WHO Guidelines on Physical Activity and Sedentary Behaviour* 2020). A definition of a healthy diet is not so easy to achieve, since nutrient intakes may vary across countries and cultures, but generally involves 55-75% of the total energy intake in carbohydrates (less than 10% free sugars), 15-30% in fats (less than 10% saturated fatty acids and less than 1% trans fatty acids), and 10-15% in proteins. Cholesterol should not exceed 300g per day. The benefit of fruits and vegetables cannot be ascribed to a mix of nutrients, and the recommendation is to consume at least 400g per day.

Regardless of the definition of multimorbidity, many studies have been published on the role of modifiable lifestyle factors on multimorbidity (Freisling et al. 2020; Dhalwani et al. 2017; Han et al. 2021; Franken et al. 2022; Aminisani et al. 2020; Geda, Janzen, and Pahwa 2021; Wikström et al. 2015; Mounce et al. 2018; Lee et al. 2022; Li et al. 2020). We found that only a few of these studies considered all five lifestyle factors (Freisling et al. 2020; Dhalwani et al. 2017; Han et al. 2021; Franken et al. 2022), and to our knowledge only one combined them in a comprehensive total lifestyle score (Freisling et al. 2020). Therefore, a gap remains regarding the role of multiple lifestyle habits combined on multimorbidity.

1.4 Aims

The aim of our study was to: i) determine the patterns of multimorbidity of specific groups of diseases or conditions using data from the Swedish National March Cohort (SNMC); and ii) estimate the effect of five important modifiable lifestyle behaviors on the incidence and multimorbidity of the selected diseases or conditions.

2 Methods

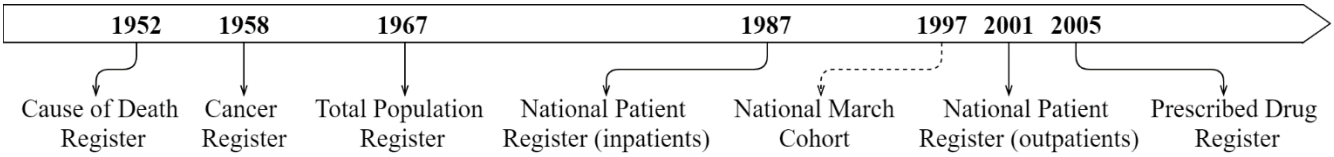
2.1 Study population

The countries of Northern Europe, and Sweden among them, have a long tradition of epidemiological research based on solid nation-wide registers. National registers may contain information either on the whole population or on a subset of it based on an experienced event or depending on certain characteristics and are established with the purpose of covering the entire target population. For this work, the members of an important Swedish cohort, the Swedish National March Cohort (SNMC) established in 1997, will be analyzed, and a virtually complete follow-up will be possible thanks to the linkage to the information contained in selected national registers.

In Sweden every resident is assigned a personal identity number (PIN) that allows linkage of records to outcomes across registers, whilst making sure that sensible information is treated with respect of privacy laws. The PIN was introduced in 1947 and consists of a concatenation of date of birth (YYMMDD or YYYYMMDD depending on the county), a three-digit birth number (odd for males, even for females), and a check digit added in 1967. A unique PIN is given to every newborn and resident staying in Sweden for at least one year. Common reasons for a PIN change are incorrect recording of date of birth or sex among immigrants or newborns (Ludvigsson et al. 2009).

Figure 2.1 shows a timeline of selected registers and the SNMC, which will be described in detail in the following paragraphs.

Figure 2.1 Timeline of the initiation of the Swedish National Registers used in this work, and establishment of the National March Cohort



2.1.1 The Swedish National March Cohort

The Swedish National March Cohort (SNMC) was established in 1997 and takes its name from a 4-day promotional event to raise funds in favour of the Swedish Cancer Society. The National March took place in almost 3600 cities and towns across the whole country, was advertised on local and national media and promoted by the Swedish King and Queen.

Participants were enrolled on a voluntary basis during the march with the expectation that they would be particularly motivated to give thoughtful answers and complete all the 36 pages of the questionnaire. Participants were encouraged to take home their questionnaire, and to return it in special mailboxes strategically situated in the stores of a large chain of supermarkets. Statistics Sweden received a total of 43,880 completed questionnaires. After the elimination of forms with an inconsistent or non-existent PIN, the cohort counted 43,863 members.

Questions were divided into various sections regarding: physical activity (the main focus of the project); diet, dietary supplements and alcohol consumption; anthropometrical measures; possible confounders including country of birth, environment during childhood and adolescence, education and employment; tobacco products consumption and passive smoke exposure; vaccines, medical and pharmacological history; sun and UV exposure and skin complexion; psychosocial history including work, life events, self-perceived health and sleep duration and quality; for women only, menstruation history, parity, infertility, use of contraceptives, menopause and use of hormonal replacement therapy.

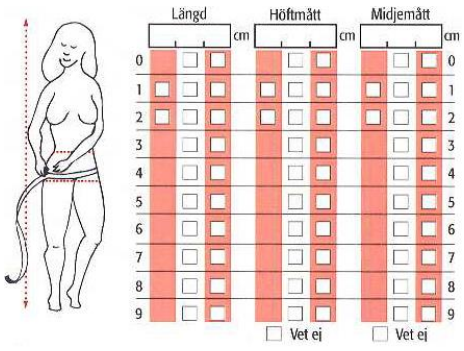
While it was not possible to estimate the proportion of respondents due to the open nature of the National March, researchers found that the proportion of missing or inconsistent answers in the returned questionnaires was exceptionally low (Trolle Lagerros et al. 2016).

2.2 Risk factors

2.2.1 *En timme för forskningen* – one hour for research

The quality of the baseline information characterizing the lifestyle of the SNMC participants widely depends also on the quality of the questionnaire presented to the volunteers. The questionnaire was advertised to take approximately one hour to complete (thus its title, "One hour for research"), and comprised 36 pages of multiple choice questions. Often, questions were accompanied by illustrations that had the purpose of making the questionnaire clearer **(Figure 2.2**, panel A) or just more fun for the users (**Figure 2.2**, panel B), as well as trivia that served the same entertainment purpose.

Figure 2.2 Examples of illustrations included in the SNMC questionnaire

A	B
	

Panel A: a woman taking her height and waist and hip circumference measures; panel B: a man sunbathing

In this paragraph I briefly report how the questionnaire was structured, and in the following how the answers were later translated into valid measures that were also comparable across other studies.

Cigarette smoking

The portion of the questionnaire relative to cigarette smoking comprised only two questions: the first asked "Have you ever smoked cigarettes daily during more than 6 months?" and had two possible answers, i.e., yes or no. The second asked "How much have you smoked in different ages?" and was followed by the possible combinations of various age classes and number of cigarettes per day in classes of width 5 (the first being 0 and the last being More than 30).

Alcohol consumption

The section regarding alcoholic beverages followed that dedicated to diet and shared a similar structure. The participant was asked "How often do you usually drink alcohol?" followed by six choices of beverages (i.e., medium strong beer, strong beer, white wine, red wine, dessert wine, and liquor), foe each of which the participant had to select one frequency option choosing between Never, 0-1 or 2-3 times per month, 1-2, 3-4, or 5-6 times per week, or 1, 2, or 3+ times per day. This question was followed by another, asking "How much do you usually drink on every occasion?". In this case there were only three possible specifications, one for beer, one for wine, and one for liquor. Each of the three types of beverage had different quantity options depending on how the beverage is usually consumed, i.e., beer could be quantified as <33cL, 33-50cL, etc. (where 33cL is a size of a small stein or can of beer, 50cL that of a medium can or stein), wine was classified in glasses and bottles, and liquor as 6cL or less, 7-12cL, 13-18cL, etc. (where 6cL is the average size of a shot of liquor).

Body composition

Participants were asked for, amongst other body measurements, their weight, height, and waist and hip circumferences. The participants were invited to write their measure in numbers and also to bar boxes corresponding to the units, tens, and hundreds of their measure (**Figure 2.2**, panel A). Waist and hip circumference had an additional box saying, "Do not know".

Physical activity

The section dedicated to physical activity included detailed questions on various kinds of activity done in free time as well as during the working day. The part of the questionnaire regarding leisure-time physical activity asked two questions, the first involving daily activities ("How much daily exercise have you got per week during the last 12 months, e.g. by walking and/or biking to work, by weekly cleaning, gardening or the alike?"), and the second involving sports ("How much time per week, on average, have you devoted to athletics/exercise/sports/outdoor life during the last 12 months?"). The first had five time classes (i.e., Less than 1 hour, 1-2, 3-4, 5-6 or More than 6 hours) as options, while the second had three possible intensities (light, such as taking a walk, moderate, such as speedy walk, jogging, or swimming, and strenuous, such as hard training and competition) each crossed with time classes (0, 0-1, 2, 3, 4, and 5 or more hours). Moreover, the latter had also two separate sets of time classes, one for summer and one for wintertime.

Diet

The dietary part of the questionnaire was one of the longest, spanning over six pages. Summarizing, food items were grouped into sections, i.e., nonalcoholic beverages (four options), sweeteners (one option), diaries (eight options), and bread (four options); fats used on bread (seven options) and for cooking (ten options); cereals (eight options); meat (excluding poultry; eight options); poultry, fish, and eggs (seven options); potatoes and carrots (four options); vegetables (twelve options); fruit and berries (seven options); other (twelve options, including sweets, biscuits, cake, chocolate, nuts, and various dressings); and fried food (five options). The frequency of consumption for each single item was normally expressed in 0 and 1-3 times per month, 1-2, 3-4 and 5-6 times per week, and 1, 2, and 3+ times per day. Only for the first section (beverages, sweeteners, diaries, and bread) the frequency was measured in glasses, cups, spoons, tablespoons, or slices, per day, depending on the item; the options touched all integers between 0 and 6, the last option being 7+.

2.2.2 Total lifestyle score

In order to put together a variety of lifestyle risk factors, we developed a multi-component score, taking as a starting point various alternatives proposed and already used on data collected within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort and all falling under the name of Healthy Lifestyle Index (HLI) (Aleksandrova et al. 2014; McKenzie et al. 2015; 2016; Naudin et al. 2019).

Scores based on these indices, proposed for the first time in 2014, combine information on smoking, alcohol intake, dietary habits, BMI, and physical activity, and were already related to colorectal, breast, gastric and pancreatic cancers, as well as to overall cancers. The HLI score is then constructed by summing the score (either binary or in quintiles) for each of the five lifestyle factors and ranges from 0 (least healthy) to either 5 or 20 (most healthy) points. From **Table 2.1** it is clear that a variety of definitions was used for each component of the HLI score, even for the same population (the EPIC cohort participants) and broadly for the same outcome (cancer at different sites).

We wanted to construct a score which was as close as possible to those already used in the framework of EPIC, but at the same time we had to come to terms with the fact that the questionnaire filled out by the SNMC participants not always allowed to calculate the HLI score components as previously proposed.

Alcohol consumption, body composition and physical activity

Grams of alcohol consumed on average per day, BMI, and METs to quantify leisure-time physical activity had already been calculated and implemented into the SNMC database (Lagerros et al. 2009; Bellocco et al. 2010).

Alcohol was estimated considering the average alcoholic content of each type of beverage crossed in the questionnaire. Where the size of the drink corresponded to more than one type of drink (e.g., both red and white wine were selected), the same size was considered for all drinks crossed by the participant.

BMI was calculated from self-reported height and weight.

METs were calculated on the basis of the Compendium of physical activities: an update of activity codes and MET intensities (Ainsworth et al. 2000).

To build our partial HLI scores, we considered increasing quintiles of METs and decreasing quintiles of alcohol and BMI. The partial HLI scores all ranged between 0 (worst lifestyle habit) and 4 (best lifestyle habit).

Table 2.1 HLI score components proposed in five studies on the EPIC cohort

Modifiable lifestyle factor	Measure unit or modality	Scores and levels of exposure	Study
Cigarette smoking	Status	1 = Never or former smoker 0 = Current smoker	A
	A combination of status, intensity, and time since cessation	4 = Never 3 = Ex-smokers > 10 years 2 = Ex-smokers ≤ 10 years 1 = Current ≤ 15 cig/day 0 = Current > 15 cig/day	B, C, D, E
Alcoholic beverages	Standard drink, equivalent to 12 g/day of alcohol	1 = Limit of two standard drinks a day for men and one for women 0 = Otherwise	A
	Alcohol consumption in g/day	4 = None 3 = 0.1-4.9 2 = 5.0-9.9 1 = 10.0-19.9 0 = ≥ 20	B
	Alcohol consumption in g/day	4 = < 6.0 3 = 6.0-11.9 2 = 12.0-23.9 1 = 24.0-59.9 0 = ≥ 60.0	C, D(McKenzie et al. 2016; Naudin et al. 2019), E
Body composition	BMI in kg/m^2 or waist circumference in cm	1 = BMI <25 kg/m^2 or waist circumference <80 cm for women and <94 cm for men 0 = Otherwise	A
	BMI in kg/m^2	Quintiles (lower is better)	B, E
	BMI in kg/m^2	4 = < 22 3 = 22-23.9 2 = 24-25.9 1 = 26-29.9 0 = ≥ 30	C
	Waist-to hip ratio	Quintiles (lower is better)	D
Physical activity	METs	1 = METs > 57 for men and METs > 82 for women 0 = METs ≤ 57 for men and METs ≤ 82 for women: 0	A
	METs	Quintiles (higher is better)	B, C, D, E
Diet	Dietary quality index including eight dietary items (fruits, vegetables, red and processed meat, fiber, fish, nuts, garlic, and yogurt)	1 = 5 to 8 points of the diet index 0 = 0 to 4 points of the diet index	A
	Intakes of seven dietary factors: cereal fiber, folate, the ratio of polyunsaturated to saturated fat, fatty fish, margarine, glycemic load, and fruits and vegetables. Consumption of dietary components were grouped into country-specific deciles and scored from 0 to 9 (inverse for trans-fat and glycemic load), with 0 being least healthy consumption.	Quintiles (higher is better)	B
	Intakes of six dietary factors: cereal fiber, red and processed meat, ratio of polyunsaturated to saturated fat, margarine, glycemic load, and fruits and vegetables. For each dietary factor, residuals were computed in models with total energy intake, and grouped into country-specific deciles.	Quintiles (higher is better)	C, D

A: (Aleksandrova et al. 2014), B: (McKenzie et al. 2015), C: (McKenzie et al. 2016), D: (Naudin et al. 2019), E (Botteri et al. 2022)

Cigarette smoking

For the HLI component of cigarette smoking, the main issue for us was the lack of precise time of cessation for former smokers. What we had, instead, was the intensity of smoking in various age classes of width 10 years, with addition of current intensity of smoking.

We came up with a definition similar to that proposed in (McKenzie et al. 2015; 2016; Naudin et al. 2019), modifying the two levels for former smokers from "quit > 10 years prior" and "quit ≤ 10 years prior" to "quit less recently", i.e., including in this level who quit in any previous age class and had age above the midpoint of current age class, and "quit more recently", i.e., including who quit in current age class, or quit in previous age class and had age below the midpoint of current age class, respectively. As an example, if a participant aged 36 years had quit in the age class 20-29, he/she would be classified as a "less recent" former smoker; conversely, if another participant aged 34 years had quit in the age class 20-29, he/she would be classified as a "more recent" former smoker. Respectively, these two participants would be attributed a score for cigarette smoking of 3 and 2.

Another issue we had to deal with was the presence of missing information in one or more of the smoking-related variables. We decided to consider as "No" every question left unanswered, if at least one other piece of information was available in the same section of the questionnaire.

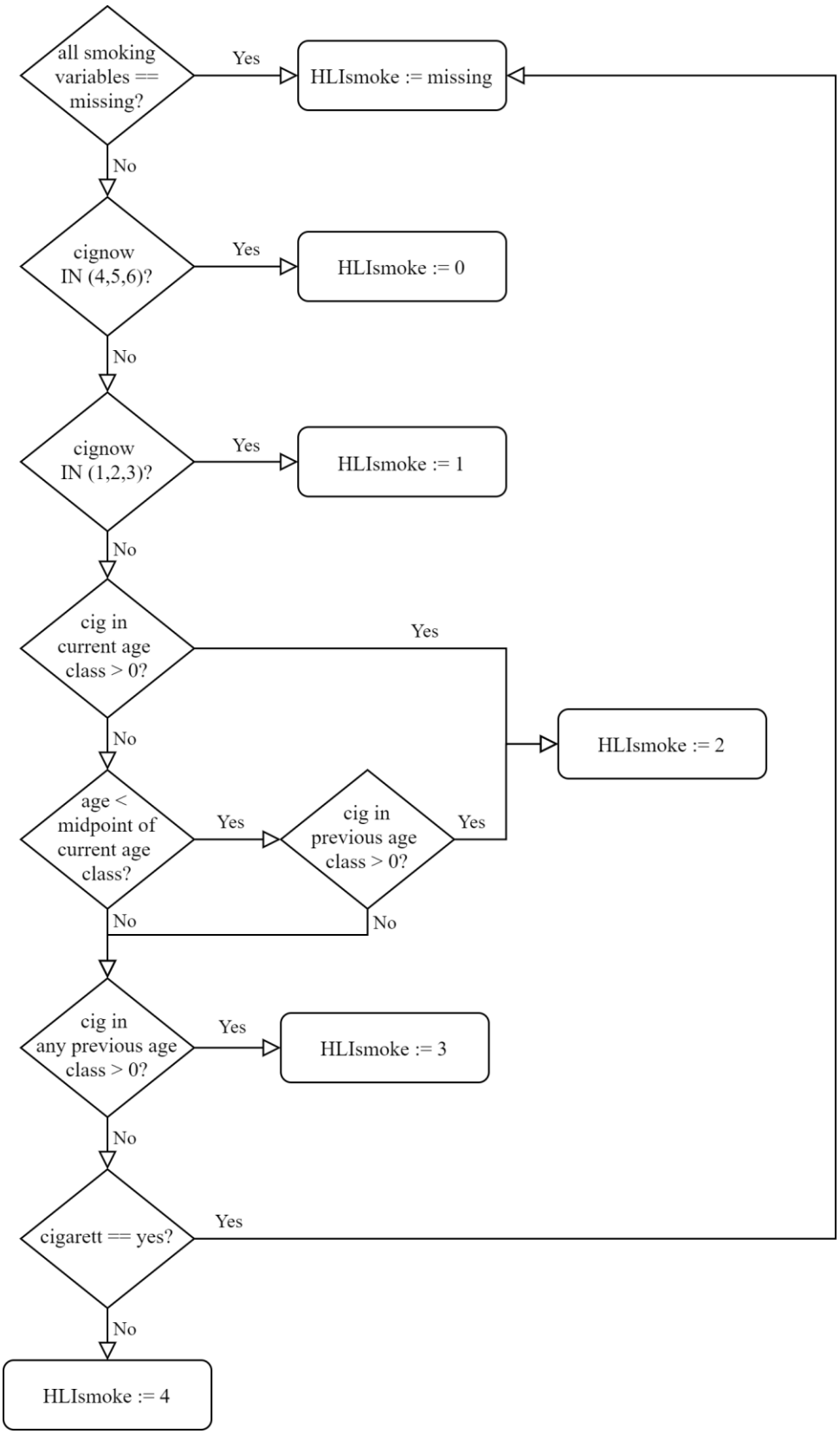
Finally, we had to beware of potentially contradictory information. The smoking-related variables found in the questionnaire can be described as follows:

- cigarett = a yes/no variable, answering to the question: "Have you ever smoked cigarettes daily for more than 6 months?"
- cignow = intensity of smoking in classes, i.e., 0 = "No", 1 = "1-5", 2 = "6-10", 3 = "11-15", 4 = "16-20", 5 = "21-30", or 6 = ">30" referring to the current intensity of smoking
- cig1014, cig1529, cig 2029, cig3039, cig4049, cig5059, cig60 = intensity in classes, i.e., 0="No", 1 = "1-5", 2 = "6-10", 3 = "11-15", 4 = "16-20", 5 = "21-30", or 6 = ">30" referring to the intensity of smoking in specific age classes. Note that smoking intensity in the current age class and cignow may have different values

A possible source of contradiction is in the cigarett variable, which identifies habitual ever smokers rather than ever smokers. Therefore, wherever possible we used the intensity variables to define the HLI score component for cigarette smoking, leaving cigarett to a final check for those classified as never smokers.

The decision diagram for the attribution of the HLI component for smoking is reported in **Figure 2.3**.

Figure 2.3 Decision diagram for the HLI score component of cigarette smoking



Diet

Due to the extensiveness of the food frequency section of the SNMC questionnaire, and to the great variety of definitions found in the literature for a healthy diet (Lassale et al. 2016), this last HLI component was the most challenging to define.

Dietary intake was assessed with an 85-item validated semi-quantitative, self-administered food frequency questionnaire (FFQ), part of the SNMC questionnaire. Participants were asked to report how many times a day, a week, or a month, on average, they consumed each food item and beverage. Linking this information to the mean portion size for each item, it was possible to obtain an estimated quantity, in grams, consumed per day by each participant.

To define the HLI score component for diet, we followed the Mediterranean Dietary Pattern (MDP) score applied in another study on a cohort of Swedish women (Yin et al. 2021). The score is defined, first, placing each food item into one among eight categories, which are: vegetables, fruits and nuts (including fruit juice), cereals, legumes, dairy products, fish and seafood, meat, or alcoholic beverages. We ignored the latter, since we already intended to include alcohol as an independent lifestyle factor. Then, a MDP score for each participant on each category is calculated, comparing the calorie-adjusted participant's consumption in grams and the median consumption in the SNMC. If the category is thought to be beneficial (i.e., vegetables, fruits and nuts, cereals, legumes, and fish and seafood), each participant with an intake greater than or equal to the cohort median is attributed a score of 1, and 0 otherwise. Instead, if the category should be consumed in moderation (i.e., dairy products, meat, and alcohol), each participant with an intake greater than or equal to the cohort median is attributed a score of 0, and 1 otherwise. Finally, one last binary variable is added to the others, based on the monounsaturated-to-saturated fat (M/S) ratio for each participant. Since monounsaturated fats should be consumed in greater amount than saturated fats, participants with values of M/S greater or equal to the median in the SNMC were attributed a score of 1, and 0 otherwise. Missing values were interpreted as null intakes (Michels and Willett 2009).

In calculating the MDP score, we did not include alcohol because we already included it as a separate component of our total HLI score, thus obtaining a MDP score ranging from 0 (worst dietary habits) to 8 (best dietary habits). We then rescaled the MDP score by dividing it into quintiles, with lower MDP scores corresponding to lower values of our HLI partial score for diet ranging from 0 to 4.

2.3 Outcomes

2.3.1 Multimorbidity definition

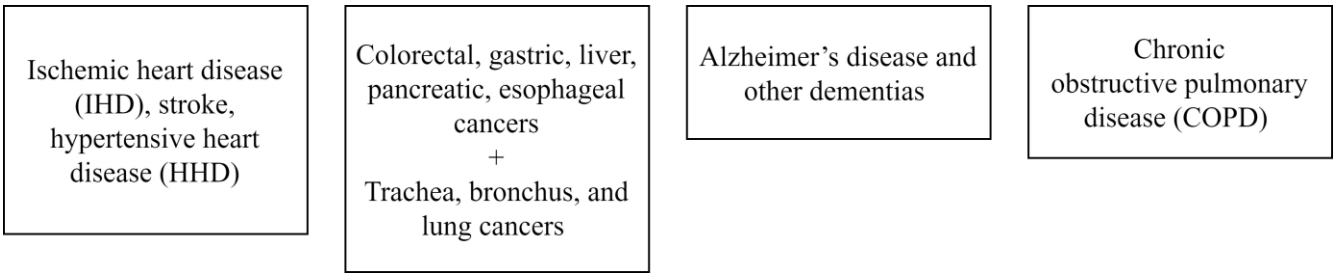
To define multimorbidity we considered all chronic causes of death among the 369 diseases, injuries, and impairments recorded in the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) (Vos et al. 2020) publicly available databases (Global Burden of Disease Collaborative Network. Seattle, United States: Institute for Health Metrics and Evaluation (IHME) 2020).

To obtain a ranking of the diseases compatible with the SNMC, we limited our search to Sweden and to calendar years between 1997 and 2016. We ranked the causes of death by decreasing yearly rates and grouped them as follows:

1) Cardiovascular diseases, i.e., ischemic heart disease (IHD; ICD-10 codes: I20-I25), stroke (I60-I69), and hypertensive heart disease (HHD; I11-I15).
2) Gastrointestinal tract (GIT) cancers, i.e., colorectal (C18-C21), gastric (C16), liver (C22), pancreatic (C25), and esophageal (C15) cancers.
3) Alzheimer's disease and other dementias (F01-F03, G30-G31).
4) Chronic obstructive pulmonary disease (COPD; J40-J44).
5) Respiratory tract (RT) cancers, i.e., trachea, bronchus, and lung cancers (C33-C34).

Subsequently we combined the two cancer groups, obtaining four macro-groups of chronic diseases (**Figure 2.4**) that are together responsible for an average cause-specific crude mortality rate of 827 deaths per thousand inhabitants. Multimorbidity was defined as the occurrence of diseases from two different groups.

Figure 2.4 Four groups of diseases



2.3.2 Assessment of the outcomes

Outcomes were assessed by linkage of the SNMC with nationwide registers, such as the cause of death register, the cancer register, and the patient register (Ludvigsson et al. 2016; 2011; Barlow et al. 2009; Brooke et al. 2017). Linkage was performed using participants' unique personal identifier. The used registers cover virtually all of Sweden, therefore providing accurate outcome assessment.

The Total Population Register

The Church of Sweden started to keep local registers of parish members as early as 1686, while the Swedish state took advantage of them for military and taxation purposes. In 1967 all local population registers were computerized, and the Total Population Register (TPR) was born. In 1991, the responsibility for local records was moved from the parishes to the local tax offices, which continued to organize data in local units. Local data are then stored nationally into the Population Register (PRTax), maintained by the Swedish Tax Agency, which in turn converges into the TPR, maintained by Statistics Sweden. Both registers share the same variables, including date and country of birth, sex, area of residence, changes of citizenship, immigration, movements within Sweden, emigration, civil status, family composition, cohabitants, and death. The TPR is updated monthly, quarterly, and annually (Ludvigsson et al. 2016).

Based on the TPR, Statistics Sweden produces every year other special registers, including the PIN register, which contains every PIN since inception, and PIN changes, had they occurred.

Close to 100% of births and deaths, 95% of immigrations and 91% of emigrations (plus at most 0.5% non-reported emigrations) are recorded into the TPR within 30 days, this proportion increasing even more over time (Ludvigsson et al. 2016).

The National Patient Register

The Swedish National Patient Register (NPR) was initiated in 1964, when the National Board of Health and Welfare started to collect information about inpatients at public hospitals, and complete national coverage was reached in 1987, when data delivery was made mandatory. Since 2001, the NPR also includes information on hospital-provided outpatient care. Within the NPR, the National Inpatient Register (IPR), also called the Hospital Discharge Register, currently holds information about more than 99% of all somatic (including surgery) and psychiatric hospital discharges, with higher data quality if the diagnosis is more severe or comprises causally-related complications, and an overall positive predictive value of diagnoses of 85% to 95% (Ludvigsson et al. 2011).

The Cancer Register

The Swedish Cancer Register was founded in 1958, and since the 1980s it is organized into six regional registers. It is compulsory for every health care provider to report new diagnoses of cancer to the register, whether the diagnosis was made clinically, by morphology, by other laboratory examinations, or at autopsy. The register covers the whole population, with under-reporting estimated at about 3.7% in 1998. Cancer diagnoses are recorded according to the current version of the ICD classification of the WHO (Barlow et al. 2009).

The Cause of Death Register

The first attempt to keep a register of deaths and specific causes such as the plague or maternal deaths dates back to 1751. Across centuries the responsibility of this register was passed from the Church to various governmental

organizations, to Statistics Sweden in 1911 when, for the first time, all causes of death were recorded. The final version of the Cause of Death register was established in 1952, moving in 1994 to the maintenance of the Swedish National Board of Health and Welfare. The Cause of Death Register is, today, a virtually complete register of all causes of death that occur among residents of Sweden. All causes of death are in compliance with the current version of the ICD classification of the WHO (Brooke et al. 2017).

The Prescribed Drug Register

The Prescribed Drug Register was founded in 2005 and is maintained by the Swedish National Board of Health and Welfare. It contains data on all prescribed drugs dispensed at pharmacies in Sweden, including the dose and price of the prescribed drug, date of prescription and date of collection at the pharmacy, and the prescriber's profession and practice. Despite only drugs prescribed for home therapy are included in the register, thus excluding therapies typically prescribed and administered in hospital such as certain antibiotics and chemotherapeutic agents, it has been estimated that the records in the Prescribed Drug Register accounted for 84% of the total utilization (Wettermark et al. 2007).

2.4 Inclusion and exclusion criteria

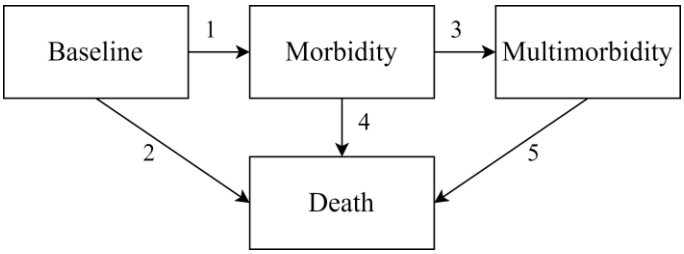
We excluded participants who had an incorrect national registration number, were aged less than 18 years, emigrated or died before time of enrollment, or immigrated (i.e., were registered) after enrollment. Moreover, we excluded participants with any prevalent malignant cancer (excluding non-melanoma skin cancer), any prevalent cardiovascular disease (ICD-10 codes: I00–I99), or any prevalent disease or condition selected for our study. Lastly, we excluded participants with missing information on smoking habits, alcohol consumption, BMI, physical activity, or diet. Follow-up started October 1st, 1997, and censoring occurred at emigration or on December 31st, 2016, whichever came first. The study was approved by the Regional Ethical Review Board at Karolinska Institutet, Stockholm, Sweden and all subjects provided informed consent.

2.5 Statistical methods

For this study, baseline variables were summarized overall and in strata of the HLI score divided in tertiles. Continuous variables were presented as means and standard deviations (SD), skewed variables as medians and interquartile ranges (IQR), and categorical variables as counts and percentages. The incidence rates of IHD, stroke, and HHD combined, GIT and RT cancers combined, dementia, COPD, multimorbidity, and death, were reported. Multi-state models (Hougaard 1999) were used to model the transitions between 1) baseline and morbidity, 2)

baseline and death, 3) morbidity and multimorbidity, 4) morbidity and death, and 5) multimorbidity and death (Figure 2.5).

Figure 2.5 Multi-state framework



We used stacked graphs to represent the probability of transitions between states of our multi-state framework at fixed values of relevant baseline covariates, and we calculated 95% confidence intervals (CIs) using 200 bootstrap samples. For each transition we tested the performances of four parametric models: exponential, Weibull, Royston-Parmar with three or four degrees of freedom. To select the best fitting model for each transition we looked for the lowest Akaike's Information Criterion (AIC). We included in the models the exposures of interest together with age, sex, and highest completed education (categorized into compulsory school or below, high school, and university or higher) as possible confounders. We reported estimates from the models which included either the continuous total HLI score, the total HLI score in five categories (0-4, 5-9, 10-12, 13-15, 16-20), or the five continuous partial HLI scores. We tested the proportional hazard assumption for each model by adding interactions of the covariates with the logarithm of the survival time (Hosmer, Lemeshow, and May 2011) and we included as time dependent covariates those with statistically significant interaction terms. We reported hazard ratios (HRs) and 95% CIs for the exposures of interest, overall and stratified by sex. A possible heterogeneity of the estimates between males and females was investigated using the Cochran's Q test. In case of time-varying effects, the estimates were reported at 5, 10, and 15 years.

As a sensitivity analysis, to identify the most influent factors among those composing the HLI score, we created five modified HLI scores ranging from 0 to 16 obtained by removing one lifestyle at a time from the calculation. The consequent models were also adjusted for the removed partial HLI score.

Statistical analyses were performed with Stata version 17.0 (College Station, TX: StataCorp LLC), R version 4.2.0 (R Core Team 2021), and SAS software version 9.4 (SAS Institute Inc).

2.5.1 Multi-state models

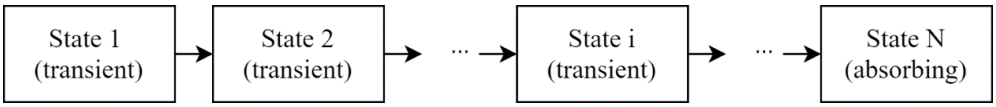
In the setting of a longitudinal study, subjects are usually followed over time until the occurrence of a particular event of interest. Sometimes, however, the condition of a patient may be split into more than one state representing the patient's life experience. Multi-state models (MSMs) have been discussed by various authors to describe the development of longitudinal data in a wide selection of epidemiological situations, while including the possible

effect of covariates on the entire process. The following paragraphs are based on some of these publications (Hougaard 1999; Commenges 1999; Andersen and Keiding 2002; Meira-Machado et al. 2009).

States can be defined on the basis of symptoms, clinical evaluations or biomarkers, while the events leading to a change of state are called transitions. A state is absorbing if cannot occur further transitions from it, while a state that is not absorbing is transient. The complexity of a MSM increases with the number of states, and depends on state structure, i.e., the states of the process and which transitions from state to state are possible.

The simplest form of MSM is the K-progressive model, illustrated in **Figure 2.6**. In a K-progressive model the subject experiences states sequentially, and the transitions are oriented in one direction only, towards the absorbing state. An example of this model is the fertility model for a woman, where the states represent the number of children, e.g., none, one, two and three or more. When two transient states and an absorbing state are present, then the model is called a progressive three-state model, e.g., the evolution of HIV infection, with states not infected, infected by HIV, and full-blown AIDS. When one transient and one absorbing state are the only possibilities, e.g., alive and dead, this results in the degenerate case of a mortality model that is handled with survival analysis techniques.

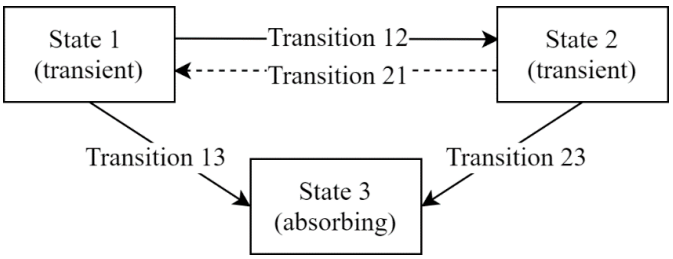
Figure 2.6 K-progressive model



When studying the evolution of a disease, another simple example of MSM is the illness-death model, illustrated in

Figure 2.7. In this example, transitions are also possible between the first and the last state, and usually backwards from the second and the first transient states. Examples of illness-death model usually involve disease-free and diseased as transient states and dead as absorbing state. When the transition from the second state back to the first state is not possible, this is also sometimes referred to as the disability model, implying the non-reversibility of the disabled intermediate state.

Figure 2.7 Illness-death model



Model specification

In general, MSMs are used to model a continuous time stochastic process, i.e., a family of random variables $\{X(t), t \in T\}$ with T an interval and values in a finite state space $S = \{1, \dots, K\}$. The observation of the process up to the time t, excluding t, is denoted with H_{t^-} and called history. The history of a multi-state process refers to all that happened through time in $[0, t)$, including the states visited and the times of transitions, and is summarized by $\{N(t), T_1, \dots, T_{N(t)}, X(0), X(T_1), \dots, X(T_{N(t)})\}$ where $N(t)$ is the number of transitions experienced until time t, and T_m with $m = 1, \dots, N(t)$ are the times of transition. The process is characterized by the state structure and the transition probabilities $p_{h,j}(s, t)$ from state h to state j, defined as

$$p_{hj}(s, t) = \text{prob}(X(t) = j \mid X(s) = h, H_{t^-})$$

with $h, j \in S$, $s, t \in T$ and $s \leq t$.

Alternatively, transition probabilities can be replaced by the transition intensities $\lambda_{hj}(t)$, which represent the instantaneous hazard of transitioning from state h to j at time t:

$$\lambda_{hj}(t) = \lim_{\Delta t \to 0} \frac{p_{hj}(t, t + \Delta t)}{\Delta t}$$

with $j \neq h$ and considering the intensity conditionally to the history H_{t^-}. We shall consider the limit to exist, and the transition intensities to be smooth and continuous. The instantaneous hazard of exiting from state h at time t is given by $\lambda_h(t) = \sum_{j \neq h} \lambda_{hj}(t)$, while that of remaining in h is assumed to be $\lambda_{hh}(t) = -\lambda_h(t)$.

Transition probabilities and intensities are collected, respectively, into a transition probabilities matrix $\mathbf{P}(s, t)$ and a transition intensities matrix $\mathbf{\Lambda}(t)$, both of dimensions $K \times K$.

Different assumptions can be added about the transition intensities and their dependence on time. A Markov model is obtained if we consider the future evolution of the process to be dependent only on the currently occupied state, i.e., $\lambda_{hj}(t)$ is independent of the history H_{t^-} for each $j \neq h$. The Markov assumption, also known as loss of memory, is not always appropriate and should be tested, for example, by including covariates depending on the history. An

even stricter assumption on the transition intensities leads to a homogenous Markov model, characterized by time-independent hazards, i.e., $\lambda_{hj}(t) = \lambda_{hj}$. Since the homogeneity assumption is too strict in many applications, non-homogenous Markov models can be simplified by partitioning time into intervals and considering transition intensities to be piecewise constant in each interval. A more general assumption consists in letting the transitions depend on the time spent in the current state but not on the history H_{t^-}, leading to semi-Markov models. Remembering that T_m is the time until the m-th transition, and $N(t)$ the number of transitions until time t, the time spent in the present state can be expressed as $t - T_{N(t^-)}$ and the semi-Markov assumption as $\lambda_{hj}(t) = \lambda_{hj}(t - T_{N(t^-)})$.

Assumptions on the population

Until this moment, we considered the same model equations for all subjects. In real applications, though, every subject i with $i = 1, \dots, M$ of the population under study is complemented with a specific set of covariates, and therefore generates a process $\{X_i(t), t \in T\}$. In order to make inference, we must assume each subject to evolve on the same states, and to experience time with the same meaning. Therefore t shall be the time passed from an event which has a particular meaning for the clinical setting, called the birth of the process. In epidemiology, t would often be the age of the subject, and the birth of the process would coincide with the birth of the individual; in clinical trials the birth of the process would often be the start of a therapy; in the case of diseases with a fast varying incidence, such as COVID-19 in spring 2020 or the HIV infection in the1980s, calendar time could be the best option.

Even with the same process structure and time definition for the entire study population, subjects may still experience different transition intensities between states due to subject-specific characteristics. Even though it may be reasonable for subgroups of a heterogeneous population, e.g., males and females, the assumption of a homogenous population is very strict. Heterogeneity is usually expressed in terms of explanatory variables $\mathbf{Z_i}$ with $i = 1, \dots, M$, and the transition intensities as $\lambda_{hj}^{i}(t) = \lambda_{hj}(t, \mathbf{Z_i})$. The study population is considered homogenous conditionally on the $\mathbf{Z_i}$ with $i = 1, \dots, M$, meaning that the subjects share the same function $\lambda_{hj}(.,.)$ for every $h, j \in S$. A substantial simplification to the model that leads to a much easier inference is the proportional hazards assumption, which asks the subject-specific intensities to be proportional to a baseline hazard, i.e., $\lambda_{hj}(t, \mathbf{Z_i}) = \lambda_{0hj}(t) \cdot f(\mathbf{Z_i})$. Moreover, the baseline hazard may remain completely unspecified as in the Cox proportional hazards model for survival analysis, or it may be assumed to be piecewise constant as in the Poisson regression models. In both cases, inference will be based on the likelihood. Sometimes the function of the covariates $f(\mathbf{Z_i})$ is assumed to take the form of an exponential, and therefore $\lambda_{hj}(t, \mathbf{Z_i}) = \lambda_{0hj} \cdot \exp(\boldsymbol{\beta}'_{\mathbf{hj}} \mathbf{Z_i})$.

Assumptions on the observations

When working with MSMs, we must come to terms with the inevitable incompleteness if observations. First, subjects are drawn from the target population, and may be prone to truncation. Moreover, the process is subdued to censoring, a problem shared by all longitudinal studies.

Truncation is manifestation of the mechanisms that lead to the inclusion of subjects into the study, especially when depending on a particular condition. Left truncation happens when subjects are selected for the study only if they did not experience a particular event for a sufficient time and are missed otherwise. Right truncation is present when the subjects are included only if they experienced a particular event within a precise period of time. The chance of a truncation must be considered when designing the study.

Censoring happens when we cannot observe on a continuous time, but rather on a finite number of distinct times, and is distinguished between three schemes: left, right, or interval censoring. Left censoring happens when we know that the event of interest happened before a particular time, but not exactly when, while in right censoring we know the state of the subject only up to a particular time. In the majority of situations, the state of a process $X_i(t)$ for an individual i is observed not on a continuous interval T, but rather in a finite number of times $t_{i0}, t_{i1}, \dots, t_{ip_i}$: this is the case of interval censoring, and describes the situation in which we do not know the exact time of transition between two states, but we know the time intervals within which the transitions occurred.

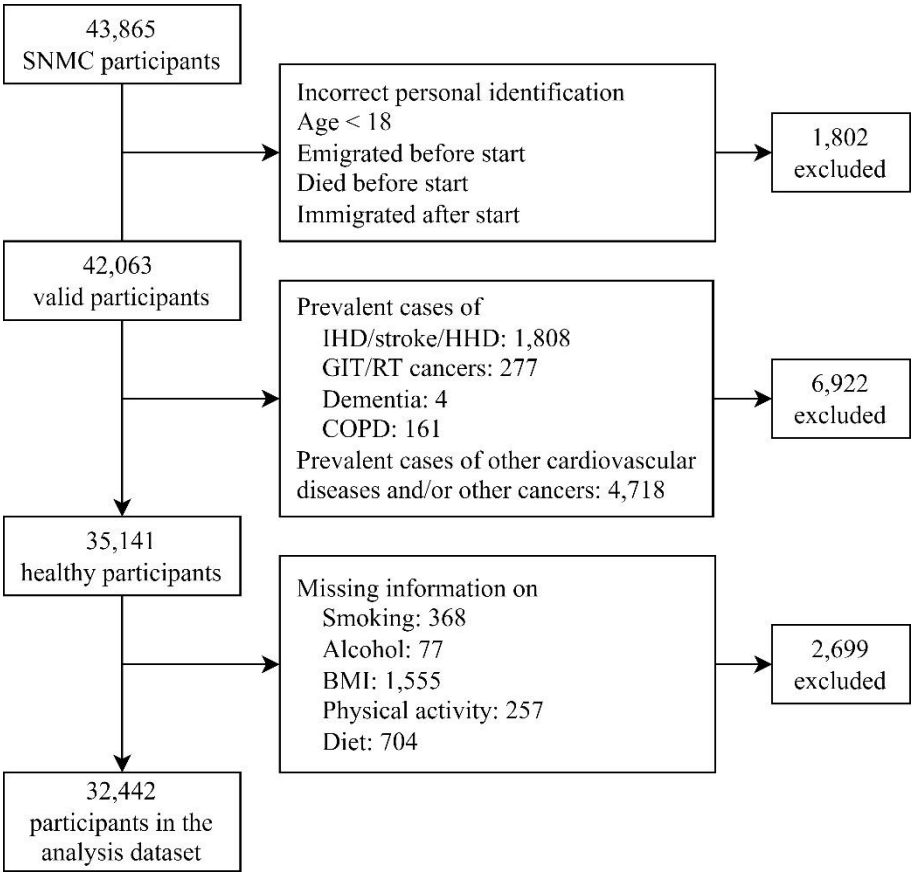
In order to write reasonably simple likelihoods, we should assume independence between the mechanism leading to incomplete observations and the process itself.

3 Results

3.1 Baseline characteristics

After applying our exclusion criteria, we included in the statistical analysis 32,442 participants without history of relevant diseases and with information on all five lifestyle exposures (**Figure 3.1**).

Figure 3.1 Cohort selection



Sex-specific demographic characteristics are reported in **Table 3.1**, overall and stratified by tertiles of the HLI score. At baseline, the included participants were aged on average 49.4 years and two-thirds were females. Around 21% of participants had completed at most compulsory education (in line with 22% in the general Swedish population in 1997, according to Statistics Sweden) and 35% held a university degree, with similar education levels between males and females. Age and education were similar by tertiles of the HLI score. The proportion of current smokers was lower compared to the 19.2% in the general Swedish population in 1997, both in males (7.0%) and females (8.4%). The average alcohol consumption was higher in males (12 grams of ethanol per day) than in females (5 g/day). The median weight (24.1 kg/m^2) for participants corresponded to a healthy BMI. The median METs indicated moderate or vigorous leisure time physical activity levels for at least half of the participants and the median MDP score for diet was 4, without noticeable differences between males and females.

Table 3.1 Selected baseline characteristics of male and female participants, overall and categorized in tertiles of the HLI score

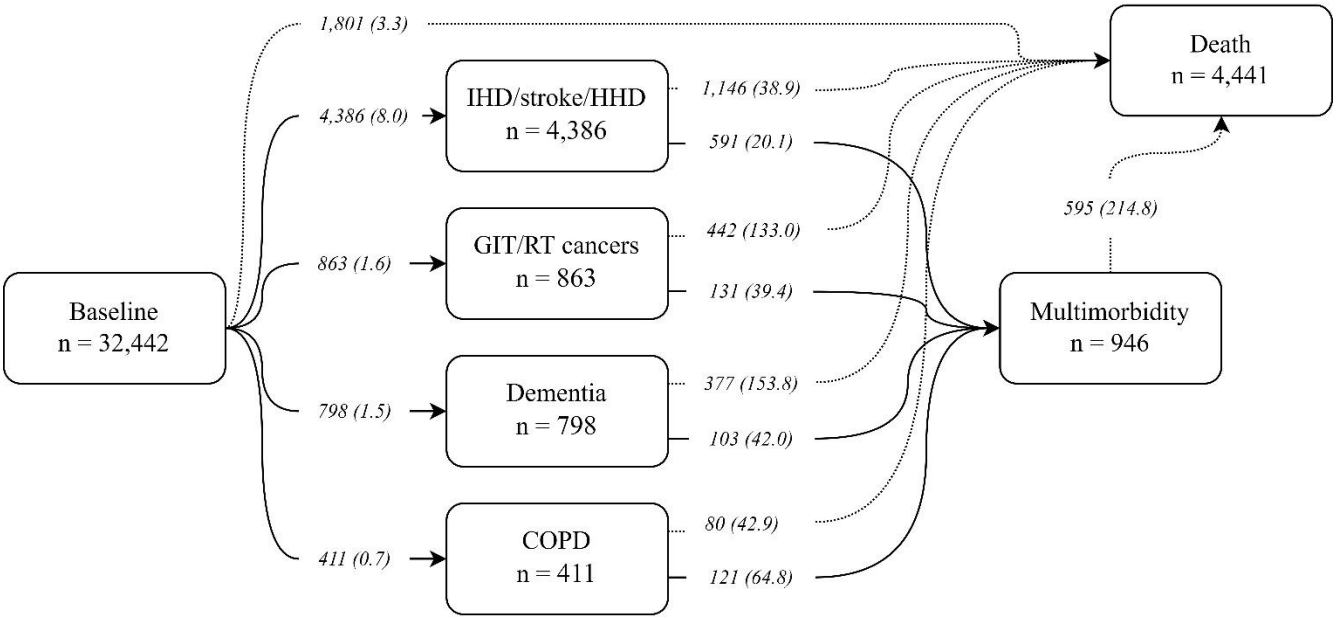
	Total HLI score in tertiles			
	HLI 0-9	**HLI 10-12**	**HLI 13-20**	**Overall**
Number of participants	9,086	11,155	12,201	32,442
Males	4,433	3,728	2,790	10,951
Age, mean (SD)	50.3 (15.1)	50.9 (16.9)	51.6 (18.7)	50.8 (16.7)
Higher education completed, n (%)				
Compulsory school or below	1,009 (22.8%)	828 (22.3%)	672 (24.2%)	2,509 (23.0%)
High school	1,996 (45.2%)	1,635 (44.0%)	1,140 (41.1%)	4,771 (43.7%)
University/PhD	1,414 (32.0%)	1,250 (33.7%)	964 (34.7%)	3,628 (33.3%)
Cigarette smoking, n (%)				
Current	648 (14.6%)	102 (2.7%)	13 (0.5%)	763 (7.0%)
Former	2,305 (52.0%)	1,369 (36.7%)	528 (18.9%)	4,202 (38.4%)
Never	1,480 (33.4%)	2,257 (60.5%)	2,249 (80.6%)	5,986 (54.7%)
Alcohol (g/day), median (IQR)	18.6 (10.2-30.9)	11.4 (4.8-20.5)	4.1 (0.6-9.6)	11.7 (4.4-22.8)
BMI (kg/m^2), median (IQR)	26.1 (24.4-28.0)	24.4 (22.9-26.0)	23.0 (21.6-24.4)	24.7 (22.9-26.6)
Phys. act. (METs), median (IQR)	3.7 (2.3-5.3)	5.7 (4.0-7.7)	7.3 (5.6-9.4)	5.3 (3.3-7.3)
MDP score, median (IQR)	3 (2-4)	4 (3-5)	5 (3-6)	4 (2-5)
Females	4,653	7,427	9,411	21,491
Age, mean (SD)	47.6 (13.4)	49.2 (14.6)	48.8 (15.9)	48.7 (14.9)
Higher education completed, n (%)				
Compulsory school or below	887 (19.1%)	1,476 (19.9%)	1,872 (20.0%)	4,235 (19.8%)
High school	2,149 (46.3%)	3,308 (44.6%)	3,958 (42.2%)	9,415 (43.9%)
University/PhD	1,608 (34.6%)	2,630 (35.5%)	3,555 (37.9%)	7,793 (36.3%)
Cigarette smoking, n (%)				
Current	1,078 (23.2%)	553 (7.5%)	179 (1.9%)	1,810 (8.4%)
Former	2,239 (48.1%)	2,905 (39.1%)	2,235 (23.8%)	7,379 (34.3%)
Never	1,336 (28.7%)	3,969 (53.4%)	6,997 (74.4%)	12,302 (57.2%)
Alcohol (g/day), median (IQR)	10.7 (4.9-20.2)	6.5 (1.8-13.1)	2.6 (0.5-7.0)	5.1 (1.1-11.7)
BMI (kg/m^2), median (IQR)	26.2 (24.0-28.7)	24.2 (22.3-26.5)	22.4 (20.9-24.1)	23.7 (21.8-26.1)
Phys. act. (METs), median (IQR)	3.3 (2.2-4.6)	4.7 (3.3-6.2)	6.4 (4.9-8.1)	5.1 (3.4-6.9)
MDP score, median (IQR)	3 (2-4)	4 (3-5)	5 (4-6)	4 (3-5)

HLI: Healthy Lifestyle Index; SD: Standard Deviation; IQR: Interquartile Range; BMI: Body Mass Index; MDP: Mediterranean Dietary Pattern; Phys. act.: physical activity.

3.2 Incidence rates and transition to multimorbidity

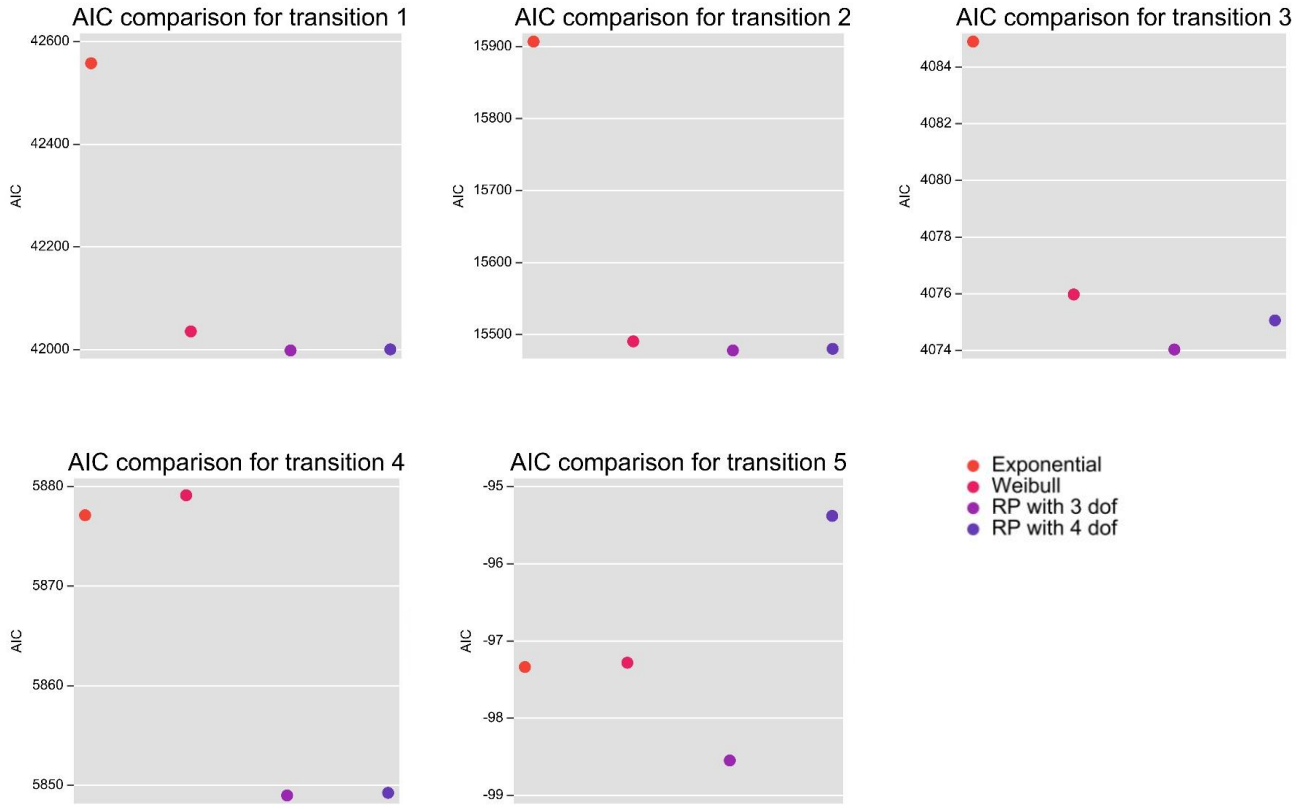
During an average follow-up time of 18.2 years the number of incident cases of one among the four selected groups of diseases was 6,458, with an average age at diagnosis of 70.3 for GIT and RT cancers, 72.1 years for COPD, 72.9 for IHD, stroke, and HHD, and 80.3 for Alzheimer's disease and other dementias. Transition to multimorbidity occurred for 946 people at an average age of 79.7 years, with the highest incidence rates for those already diagnosed with COPD (64.8 per 1,000 person-years), and the lowest for those diagnosed with a cardiovascular condition first (20.1 per 1,000 person-years) (**Figure 3.2**). In total we observed 4,441 deaths. For each transition a Royston-Parmar parametric models with three degrees of freedom was selected (**Figure 3.3**).

Figure 3.2 Transitions between baseline, the four diseases groups considered separately, multimorbidity, and death



Between brackets: incidence rates per 1,000 person-years. IHD: ischemic heart disease; HHD: hypertensive heart disease; GIT: gastrointestinal tract; RT: respiratory tract; COPD: chronic obstructive pulmonary disease.

Figure 3.3 AIC comparison for model selection



3.3 HLI score

We found that the partial continuous HLI components performed better in terms of goodness of fit compared with the total HLI score, either continuous or categorical.

A one-unit increase in the partial HLI score for smoking was associated with a decreased risk (HR [95% CI]) of transitioning to morbidity of 0.81 (0.79-0.83) (**Table 3.2**). Higher scores for smoking also decreased the risk of further transitioning to multimorbidity, with a stronger effect in females (for +1 unit: 0.71 [0.65-0.77]) than in males (0.79 [0.72-0.87]). Consuming more alcohol was associated with lower risk of morbidity (for +1 unit of the partial HLI score for alcohol: 1.05 [1.03-1.07]) but it was not statistically significantly associated with multimorbidity, and neither were physical activity and BMI. The positive effect of physical activity on morbidity decreased over time, with an HR (95% CI) at 5 years of 0.90 (0.86-0.94) for one unit increase in the partial HLI score, 0.95 (0.91-0.98) at 10 years and 0.98 (0.94-1.02) at 15 years. The BMI score was also positively associated

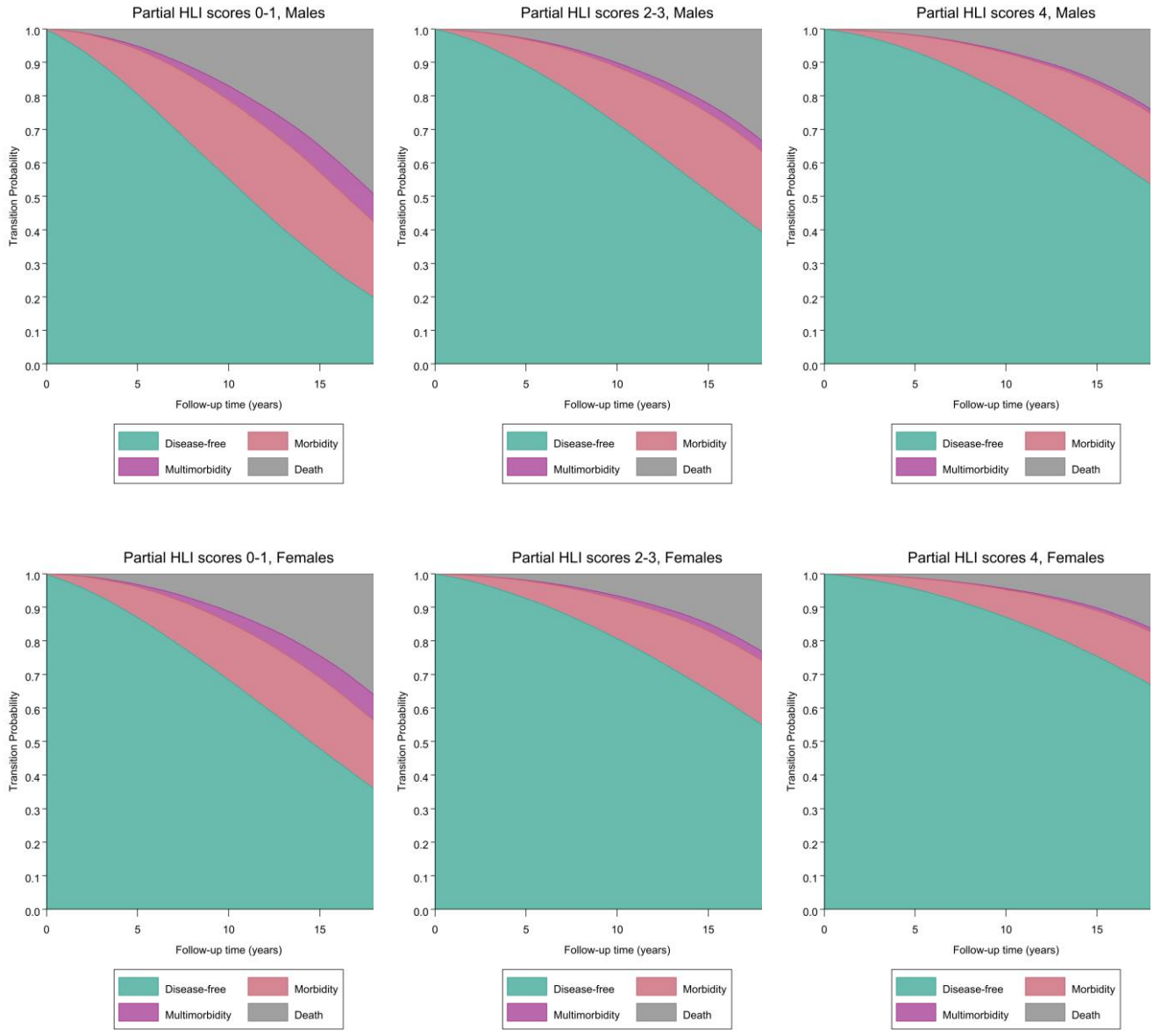
Table 3.2 Association between HLI score and the probability of transitioning from baseline to morbidity and from morbidity to multimorbidity, overall and stratified by sex

	From baseline to morbidity				From morbidity to multimorbidity			
	Overall	**Males**	**Females**	***p****	**Overall**	**Males**	**Females**	***p****
Model	**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**		**HR (95% CI)**	**HR (95% CI)**	**HR (95% CI)**	
Continuous HLI score								
+1 unit	0.96 (0.95-0.97)	0.95 (0.94-0.96)	0.96 (0.95-0.97)	*0.21*	0.94 (0.92-0.96)	0.94 (0.91-0.97)	0.95 (0.92-0.98)	*0.49*
HLI score in five categories								
HLI 5-9 vs. 0-4	0.72 (0.61-0.86)	0.77 (0.63-0.95)	0.63 (0.46-0.87)	*0.01*	0.47 (0.34-0.67)	0.48 (0.31-0.73)	0.42 (0.23-0.76)	*<0.01*
HLI 10-12 vs. 0-4	0.62 (0.52-0.74)	0.69 (0.56-0.85)	0.53 (0.39-0.74)		0.33 (0.23-0.46)	0.38 (0.24-0.58)	0.25 (0.14-0.45)	
HLI 13-15 vs. 0-4	0.56 (0.47-0.67)	0.59 (0.48-0.74)	0.49 (0.36-0.67)		0.36 (0.25-0.51)	0.36 (0.23-0.57)	0.31 (0.17-0.55)	
HLI 16-20 vs. 0-4	0.51 (0.42-0.61)	0.47 (0.36-0.61)	0.46 (0.33-0.64)		0.35 (0.24-0.52)	0.35 (0.20-0.63)	0.30 (0.16-0.56)	
Partial HLI score components								
Smoke (+1 unit)	0.81 (0.79-0.83)	0.80 (0.77-0.83)	0.81 (0.79-0.84)	*0.31*	0.74 (0.69-0.78)	0.79 (0.72-0.87)	0.71 (0.65-0.77)	*0.02*
Alcohol (+1 unit)	1.05 (1.03-1.07)	1.05 (1.02-1.08)	1.05 (1.03-1.08)	*0.78*	0.97 (0.92-1.02)	0.97 (0.90-1.04)	0.97 (0.91-1.04)	*0.92*
Phys. act. (+1 unit)					0.97 (0.93-1.02)	0.94 (0.88-1.01)	0.99 (0.93-1.06)	*0.26*
at 5 years	0.90 (0.86-0.94)	0.90 (0.85-0.96)	0.90 (0.84-0.96)	*0.93*				
at 10 years	0.95 (0.91-0.98)	0.97 (0.92-1.02)	0.94 (0.89-0.99)	*0.41*				
at 15 years	0.98 (0.94-1.02)	1.00 (0.94-1.07)	0.96 (0.91-1.02)	*0.32*				
BMI (+1 unit)	0.92 (0.91-0.94)	0.89 (0.87-0.92)	0.94 (0.92-0.97)	*<0.01*	1.03 (0.98-1.08)	1.03 (0.95-1.11)	1.03 (0.97-1.10)	*0.93*
Diet (+1 unit)	1.00 (0.98-1.02)	1.01 (0.98-1.03)	1.00 (0.98-1.03)	*0.83*	0.96 (0.91-1.00)	0.93 (0.86-1.00)	0.98 (0.92-1.05)	*0.25*

HLI: Healthy Lifestyle Index; HR: hazard ratio; CI: confidence interval; BMI: body mass index; Phys. act.: physical activity. All models are adjusted for age, sex, and education. *: p-value for the heterogeneity of the estimates between males and females.

with morbidity (for +1 unit: 0.92 [0.91-0.94]), with a stronger effect in males (0.89 [0.87-0.92]) than females (0.94 [0.92-0.97]) (p-value for heterogeneity: 0.002). Finally, the diet score seemed to have no effect on morbidity, whereas a healthier diet seemed to be protective against multimorbidity (HR [95% CI] for +1 unit of the diet HLI: 0.96 [0.91-1.00]).In general, for values of the partial HLI scores corresponding to healthier lifestyles we observed a reduction in the risk of morbidity (difference in probability [95% CI] for partial HLI scores 4 vs. 0-1 estimated at 15 years, for males: -0.02 [-0.04; 0.01], and for females: -0.07 [-0.10; -0.04]), multimorbidity (males: -0.08 [-0.10; -0.05], females: -0.06 [-0.08; -0.04]), and mortality (males: -0.20 [-0.23; -0.16], females: -0.15 [-0.18; -0.13]) (**Figure 3.4**, **Table 3.3**).

Figure 3.4 Sex-specific stacked plot of the transition probabilities for participants aged 65 at baseline by increasing values of all partial HLI scores



One unit increase in the continuous HLI score corresponded to 4% reduction in the risk of morbidity (HR [95% CI]: 0.96 [0.95-0.97]) and 6% in the risk of multimorbidity (0.94 [0.92-0.96]), similarly for both males and females (**Table 3.2**). Having an HLI score between 16 and 20 halved the risk of morbidity compared to an HLI between 0 and 4 and reduced by almost two thirds the risk of multimorbidity. In both cases, the effect was slightly stronger for females.

Table 3.3 Sex-specific transition probabilities for participants aged 65 at baseline, by increasing values of all partial HLI scores

			Partial HLI scores 0-1	Partial HLI scores 2-3	Partial HLI scores 4
		Time (years)	Prob. (95% CI)	Prob. (95% CI)	Prob. (95% CI)
Males	**Morbidity**	5	0.14 (0.13-0.16)	0.08 (0.07-0.09)	0.05 (0.04-0.06)
		10	0.24 (0.22-0.26)	0.17 (0.15-0.18)	0.12 (0.11-0.13)
		15	0.25 (0.22-0.28)	0.23 (0.22-0.25)	0.19 (0.18-0.21)
	Multimorbidity	5	0.01 (0.00-0.01)	0.00 (0.00-0.00)	0.00 (0.00-0.00)
		10	0.04 (0.03-0.06)	0.01 (0.01-0.01)	0.00 (0.00-0.01)
		15	0.08 (0.06-0.11)	0.02 (0.02-0.03)	0.01 (0.01-0.01)
	Death	5	0.05 (0.05-0.06)	0.03 (0.02-0.03)	0.02 (0.02-0.03)
		10	0.18 (0.16-0.20)	0.10 (0.09-0.11)	0.07 (0.06-0.08)
		15	0.36 (0.32-0.39)	0.22 (0.20-0.24)	0.16 (0.14-0.18)
	Disease-free survival	5	0.79 (0.77-0.81)	0.89 (0.88-0.90)	0.93 (0.92-0.94)
		10	0.54 (0.51-0.57)	0.72 (0.71-0.74)	0.80 (0.79-0.82)
		15	0.31 (0.28-0.34)	0.53 (0.5-0.55)	0.64 (0.62-0.67)
Females	**Morbidity**	5	0.09 (0.08-0.10)	0.05 (0.05-0.06)	0.03 (0.03-0.04)
		10	0.17 (0.15-0.19)	0.12 (0.11-0.13)	0.08 (0.07-0.09)
		15	0.20 (0.18-0.23)	0.18 (0.17-0.19)	0.13 (0.12-0.15)
	Multimorbidity	5	0.01 (0.00-0.01)	0.00 (0.00-0.00)	0.00 (0.00-0.00)
		10	0.03 (0.02-0.04)	0.01 (0.00-0.01)	0.00 (0.00-0.00)
		15	0.07 (0.05-0.09)	0.02 (0.02-0.03)	0.01 (0.01-0.01)
	Death	5	0.04 (0.03-0.04)	0.02 (0.02-0.02)	0.01 (0.01-0.01)
		10	0.12 (0.10-0.13)	0.06 (0.06-0.07)	0.04 (0.04-0.05)
		15	0.25 (0.23-0.28)	0.14 (0.13-0.16)	0.10 (0.09-0.11)
	Disease-free survival	5	0.86 (0.04-0.85)	0.93 (0.02-0.92)	0.96 (0.01-0.95)
		10	0.68 (0.12-0.65)	0.81 (0.06-0.80)	0.87 (0.04-0.86)
		15	0.47 (0.25-0.44)	0.66 (0.14-0.64)	0.76 (0.10-0.74)

HLI: Healthy Lifestyle Index; CI: confidence interval; Prob.: probability

3.3.1 Sensitivity analysis

As a sensitivity analysis, we removed one component at a time from the total HLI score (**Table 3.4**). We found that smoking was the most influent factor for both morbidity and multimorbidity. Smoking was followed by physical activity and BMI for morbidity, while on multimorbidity physical activity had no effect, and BMI had an inverse effect. Increasing alcohol consumption showed a protective effect on morbidity, whereas on multimorbidity the effect was detrimental. Diet did not seem to be significantly influencing the effect of the HLI score.

Table 3.4 Sensitivity analysis on the association between the continuous HLI score and the probability of transitioning from baseline to morbidity and from morbidity to multimorbidity, removing one component

	From baseline to morbidity	From morbidity to multimorbidity
Model	**HR (95% CI)**	**HR (95% CI)**
Reference		
+1 unit in the HLI score	0.96 (0.95-0.97)	0.94 (0.92-0.96)
Removing smoking		
+1 unit in the modified HLI score	0.98 (0.97-0.99)	0.98 (0.96-1.01)
+1 unit in the smoking score	0.82 (0.80-0.85)	0.73 (0.69-0.78)
Removing alcohol		
+1 unit in the modified HLI score	0.94 (0.93-0.95)	0.95 (0.92-0.97)
+1 unit in the alcohol score	1.02 (1.00-1.04)	0.93 (0.89-0.98)
Removing physical activity		
+1 unit in the modified HLI score	0.96 (0.95-0.97)	0.93 (0.91-0.95)
+1 unit in the phys. act. score		0.99 (0.94-1.03)
at 5 years	0.94 (0.92-0.96)	
at 10 years	0.96 (0.95-0.98)	
at 15 years	0.98 (0.96-1.00)	
Removing BMI		
+1 unit in the modified HLI score	0.97 (0.96-0.98)	0.92 (0.90-0.94)
+1 unit in the BMI score	0.92 (0.90-0.94)	1.05 (1.00-1.10)
Removing diet		
+1 unit in the modified HLI score	0.95 (0.94-0.96)	0.94 (0.92-0.97)
+1 unit in the diet score	0.99 (0.98-1.01)	0.95 (0.90-0.99)

HLI: Healthy Lifestyle Index; HR: hazard ratio; CI: confidence interval; BMI: body mass index; phys. act.: physical activity. All models are adjusted for age, sex, education, and the removed partial HLI score.

4 Discussion

In this large cohort of 43,865 adult participants, we found that healthy lifestyle habits, summarized by the HLI score, were inversely associated with morbidity and multimorbidity of cardiovascular diseases, gastrointestinal and respiratory cancers, dementia, and COPD. Smoking was the risk factor most strongly associated with both progression towards morbidity and to multimorbidity. High BMI and low levels of physical activity were associated with a higher risk of morbidity. High alcohol consumption was associated with a higher risk of multimorbidity, but a lower risk of morbidity. Adherence to the Mediterranean Dietary Pattern was associated with a decreased risk of multimorbidity.

Public health implications

Despite many studies having already been published on lifestyle and multimorbidity, the cumulative effect of five relevant lifestyle habits on morbidity, multimorbidity and death has remained unclear. When we stratified our study population by lower, medium, and higher partial HLI scores, we observed a reduction in the cumulative probabilities of developing the outcomes as the scores increase. For example, over 15 years of follow-up, a man aged 65 years at baseline with an excellent lifestyle (all partial scores 4) would have a 33% reduction in the cumulative probability of morbidity, multimorbidity, and death combined compared to another man with same characteristics but a poor lifestyle (all partial scores 0-1). In case of women, the cumulative probability would be reduced by 29%.

When we analyzed the total HLI score, we found that it was strongly associated with morbidity and multimorbidity, each increasing point in the score reducing the risk of morbidity by 4% and the risk of multimorbidity by 6%. Moreover, when categorizing the total HLI score in five categories, a linear relationship described the effect of this variable on morbidity well. However, this was not true for multimorbidity, where we observed a similar effect of the last three categories compared to the reference.

Our results in the context of previous studies

Our results add on to several other studies on lifestyle and multimorbidity, although direct comparison is complicated by widely differing definitions for multimorbidity (Willadsen et al. 2016). We defined multimorbidity as only chronic diseases with a high probability of reducing life expectancy (Global Burden of Disease Collaborative Network. Seattle, United States: Institute for Health Metrics and Evaluation (IHME) 2020), while previous studies used different rationales to define multimorbidity: some authors chose an arbitrary and limited set of diseases (Freisling et al. 2020; Han et al. 2021; Li et al. 2020); others allowed a broad definition of multimorbidity (either by choosing a numerous set of chronic diseases or using chronic medications as proxy of chronic diseases (Dhalwani et al. 2017; Franken et al. 2022; Aminisani et al. 2020; Geda, Janzen, and Pahwa 2021; Mounce et al. 2018; Lee et al. 2022)); and one selected the five most frequent chronic conditions in the population (Wikström et al. 2015). While our conservative definition of multimorbidity allowed us to establish the role of

lifestyle factors in preventing the occurrence of clinical outcomes in a well-defined group of patients, comparability of our study with others is reduced.

Our findings support previous evidence that smoking, low physical activity levels, and high BMI are directly associated with multimorbidity. While for smoking, physical activity, and BMI similar findings were reported consistently across studies, different results have been reported on the direction and strength of the association between alcohol consumption and multimorbidity. In a study on a Swedish population-based cohort of people aged 60 years or more (SNAC-K), the authors assessed how alcohol consumption, among other risk factors has different distributions depending on the cluster of multimorbidity considered (Marengoni et al. 2020). This can partially justify why alcohol has been found to be either inversely (Aminisani et al. 2020; Geda, Janzen, and Pahwa 2021) or positively associated with the risk of multimorbidity (Lee et al. 2022; Li et al. 2020), depending on the definition of multimorbidity used in the study.

Moreover, a few studies (Dhalwani et al. 2017; Han et al. 2021) assumed a U-shaped effect for alcohol exposure, collapsing the categories of no alcohol consumption and heaviest alcohol use (Marmot et al. 1981). Whether or not to study the effect of alcohol on a quadratic scale has been subject of discussion in the scientific community since the 1980s ("Alcohol and Mortality: The Myth of the U-Shaped Curve" 1988). The idea that no alcohol consumption might be detrimental towards morbidity (in particular towards cardiovascular disease) and overall mortality has been losing traction due to the argument that abstainers are often former heavy drinkers rather than lifelong nondrinkers, and their lifestyle included habits that put them at a higher risk of morbidity and mortality ("Alcohol and Mortality: The Myth of the U-Shaped Curve" 1988). A recent study further disproved the idea, stressing that "any reduction in alcohol consumption is in fact beneficial in terms of general health" (Chudzińska et al. 2022).

Diet was the lifestyle factor least considered in the literature and, when studied, remained inconclusive (Freisling et al. 2020; Wikström et al. 2015); we observed a possible inverse relationship with multimorbidity.

Strengths and limitations

Our study is strengthened by almost 20 years of follow-up and the linkage of our cohort of 43,865 adult participants to the virtually complete Swedish national registers (Ludvigsson et al. 2011). We considered five lifestyle exposures jointly, measured with self-reported but high quality information (Trolle Lagerros et al. 2016), and studied chronic diseases that are known to lead to death. Additionally, the use of multistate models gave us insights on how our composite lifestyle exposure might affect clinical progression between different states.

Among limitations, having participants on average healthier than the general Swedish population could lead to biased estimates of the associations between exposures and outcomes. Such bias could be in either direction, but is believed to be weak (Pizzi et al. 2011). Exposures were self-reported, leading to possible misclassification, although misclassification should be non-differential due to the prospective design. Exposures were also only

measured once (at baseline) and therefore we could not account for lifestyle changes during the follow-up. Lastly, as the nature of this study is observational, residual confounding cannot be excluded.

Conclusions

To conclude, we determined that being a never smoker or quitting smoking, having a low alcohol consumption, high physical activity levels, a low BMI, and following the Mediterranean Diet recommendations can lower the probability of morbidity and multimorbidity of death-accelerating diseases and death. This effect is particularly evident when all the healthy lifestyles are combined.

5 Other works on lifestyle

During my last year of PhD, I continued the collaboration, already active during my second year, with the International Agency for Research on Cancer (IARC). The purpose of the research was to study another important aspect of lifestyle on human health, i.e., lifestyle changes.

Using data from the EPIC cohort, we could determine the trajectories of lifestyle between two questionnaires, administered to participants at two timepoints, and the impact of changes in lifestyle on the risk of malignant cancer at any site. In particular, we first conducted a pilot study on colorectal cancer, and then we progressed to analyze all cancers. As a third but not less important outcome, all-cause mortality was also considered.

These three studies lead to novel and important results highlighting not only the necessity of conducting a healthy lifestyle from a young age, but that an improvement in lifestyle later in life still brings benefits in terms of all named outcomes.

The first study on colorectal cancer, entitled “Changes in lifestyle and risk of colorectal cancer in the European Prospective Investigation into Cancer and Nutrition” has recently been published on the American Journal of Gastroenterology (Botteri et al. 2022), while the other two are currently in writing.

6 References

Ainsworth, Barbara E., William L. Haskell, Melicia C. Whitt, Melinda L. Irwin, Ann M. Swartz, Scott J. Strath, William L. O'Brien, et al. 2000. "Compendium of Physical Activities: An Update of Activity Codes and MET Intensities:" *Medicine & Science in Sports & Exercise* 32 (Supplement): S498–516. https://doi.org/10.1097/00005768-200009001-00009.

Akker, Marjan van den, Frank Buntinx, and J André Knottnerus. 1996. "Comorbidity or Multimorbidity - What's in a Name? A Review of the Literature." *European Journal of General Practice* 2 (2): 65–70. https://doi.org/10.3109/13814789609162146.

"Alcohol and Mortality: The Myth of the U-Shaped Curve." 1988. *Lancet (London, England)* 2 (8623): 1292–93.

Aleksandrova, Krasimira, Tobias Pischon, Mazda Jenab, H Bas Bueno-de-Mesquita, Veronika Fedirko, Teresa Norat, Dora Romaguera, et al. 2014. "Combined Impact of Healthy Lifestyle Factors on Colorectal Cancer: A Large European Cohort Study." *BMC Medicine* 12 (1): 168. https://doi.org/10.1186/s12916-014-0168-4.

Almirall, José, and Martin Fortin. 2013. "The Coexistence of Terms to Describe the Presence of Multiple Concurrent Diseases." *Journal of Comorbidity* 3: 4–9. https://doi.org/10.15256/joc.2013.3.22.

Aminisani, Nayyereh, Christine Stephens, Joanne Allen, Fiona Alpass, and Seyed Morteza Shamshirgaran. 2020. "Socio-Demographic and Lifestyle Factors Associated with Multimorbidity in New Zealand." *Epidemiology and Health* 42: e2020001. https://doi.org/10.4178/epih.e2020001.

Andersen, Per Kragh, and Niels Keiding. 2002. "Multi-State Models for Event History Analysis." *Statistical Methods in Medical Research* 11 (2): 91–115. https://doi.org/10.1191/0962280202SM276ra.

Barlow, Lotti, Kerstin Westergren, Lars Holmberg, and Mats Talbäck. 2009. "The Completeness of the Swedish Cancer Register: A Sample Survey for Year 1998." *Acta Oncologica (Stockholm, Sweden)* 48 (1): 27–33. https://doi.org/10.1080/02841860802247664.

Barnett, Karen, Stewart W Mercer, Michael Norbury, Graham Watt, Sally Wyke, and Bruce Guthrie. 2012. "Epidemiology of Multimorbidity and Implications for Health Care, Research, and Medical Education: A Cross-Sectional Study." *The Lancet* 380 (9836): 37–43. https://doi.org/10.1016/S0140-6736(12)60240-2.

Bellocco, Rino, Chongqi Jia, Weimin Ye, and Ylva Trolle Lagerros. 2010. "Effects of Physical Activity, Body Mass Index, Waist-to-Hip Ratio and Waist Circumference on Total Mortality Risk in the Swedish National March Cohort." *European Journal of Epidemiology* 25 (11): 777–88. https://doi.org/10.1007/s10654-010-9497-6.

Bennett, James E, Gretchen A Stevens, Colin D Mathers, Ruth Bonita, Jürgen Rehm, Margaret E Kruk, Leanne M Riley, et al. 2018. "NCD Countdown 2030: Worldwide Trends in Non-Communicable Disease Mortality and Progress towards Sustainable Development Goal Target 3.4." *The Lancet* 392 (10152): 1072–88. https://doi.org/10.1016/S0140-6736(18)31992-5.

Botteri, Edoardo, Giulia Peveri, Paula Berstad, Vincenzo Bagnardi, Sairah L.F. Chen, Torkjel M. Sandanger, Geir Hoff, et al. 2022. "Changes in Lifestyle and Risk of Colorectal Cancer in the European Prospective Investigation Into Cancer and Nutrition." *American Journal of Gastroenterology* Publish Ahead of Print (December). https://doi.org/10.14309/ajg.0000000000002065.

Brooke, Hannah Louise, Mats Talbäck, Jesper Hörnblad, Lars Age Johansson, Jonas Filip Ludvigsson, Henrik Druid, Maria Feychting, and Rickard Ljung. 2017. "The Swedish Cause of Death Register." *European Journal of Epidemiology* 32 (9): 765–73. https://doi.org/10.1007/s10654-017-0316-1.

Busse, Reinhard, Miriam Blümel, David Scheller-Kreinsen, and Annette Zentner, eds. 2010. *Tackling Chronic Disease in Europe: Strategies, Interventions and Challenges*. Observatory Studies Series, no. 20. Copenhagen: World Health Organization on behalf of the European Observatory on Health Systems and Policies.

Chudzińska, Małgorzata, Łukasz Wołowiec, Joanna Banach, Daniel Rogowicz, and Grzegorz Grześk. 2022. "Alcohol and Cardiovascular Diseases—Do the Consumption Pattern and Dose Make the Difference?" *Journal of Cardiovascular Development and Disease* 9 (10): 317. https://doi.org/10.3390/jcdd9100317.

Commenges, D. 1999. "Multi-State Models in Epidemiology." *Lifetime Data Analysis* 5: 315–27.

Deeks, Steven G, Sharon R Lewin, and Diane V Havlir. 2013. "The End of AIDS: HIV Infection as a Chronic Disease." *The Lancet* 382 (9903): 1525–33. https://doi.org/10.1016/S0140-6736(13)61809-7.

Dhalwani, Nafeesa N., Francesco Zaccardi, Gary O'Donovan, Patrice Carter, Mark Hamer, Thomas Yates, Melanie Davies, and Kamlesh Khunti. 2017. "Association Between Lifestyle Factors and the Incidence of Multimorbidity in an Older English Population." *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 72 (4): 528–34. https://doi.org/10.1093/gerona/glw146.

European Commission. 2019. "Eurostat Databases." 2019. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=proj_19np&lang=en.

Feinstein, A. R. 1970. "The Pre-Therapeutic Classification of Co-Morbidity in Chronic Disease." *Journal of Chronic Diseases* 23 (7): 455–68. https://doi.org/10.1016/0021-9681(70)90054-8.

Franken, Débora Luiza, Maria Teresa Anselmo Olinto, Juvenal Soares Dias-da-Costa, Fernanda Souza de Bairros, and Vera Maria Vieira Paniz. 2022. "Lifestyle, High Body Mass Index, and Markers of Socioeconomic Conditions Associated with Multimorbidity in Women." *Revista Brasileira De Epidemiologia = Brazilian Journal of Epidemiology* 25: e220007. https://doi.org/10.1590/1980-549720220007.

Freisling, Heinz, Vivian Viallon, Hannah Lennon, Vincenzo Bagnardi, Cristian Ricci, Adam S. Butterworth, Michael Sweeting, et al. 2020. "Lifestyle Factors and Risk of Multimorbidity of Cancer and Cardiometabolic Diseases: A Multinational Cohort Study." *BMC Medicine* 18 (1): 5. https://doi.org/10.1186/s12916-019-1474-7.

Gakidou, Emmanuela, Ashkan Afshin, Amanuel Alemu Abajobir, Kalkidan Hassen Abate, Cristiana Abbafati, Kaja M Abbas, Foad Abd-Allah, et al. 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." *The Lancet* 390 (10100): 1345–1422. https://doi.org/10.1016/S0140-6736(17)32366-8.

Geda, Nigatu Regassa, Bonnie Janzen, and Punam Pahwa. 2021. "Chronic Disease Multimorbidity among the Canadian Population: Prevalence and Associated Lifestyle Factors." *Archives of Public Health = Archives Belges De Sante Publique* 79 (1): 60. https://doi.org/10.1186/s13690-021-00583-7.

Global Burden of Disease Collaborative Network. Seattle, United States: Institute for Health Metrics and Evaluation (IHME). 2020. "Global Burden of Disease Study 2019 (GBD 2019) Results." 2020. https://vizhub.healthdata.org/gbd-results/.

Gnjidic, Danijela, Sarah N. Hilmer, Fiona M. Blyth, Vasi Naganathan, Louise Waite, Markus J. Seibel, Andrew J. McLachlan, Robert G. Cumming, David J. Handelsman, and David G. Le Couteur. 2012. "Polypharmacy Cutoff and Outcomes: Five or More Medicines Were Used to Identify Community-Dwelling Older Men at Risk of Different Adverse Outcomes." *Journal of Clinical Epidemiology* 65 (9): 989–95. https://doi.org/10.1016/j.jclinepi.2012.02.018.

Han, Yuting, Yizhen Hu, Canqing Yu, Yu Guo, Pei Pei, Ling Yang, Yiping Chen, et al. 2021. "Lifestyle, Cardiometabolic Disease, and Multimorbidity in a Prospective Chinese Study." *European Heart Journal* 42 (34): 3374–84. https://doi.org/10.1093/eurheartj/ehab413.

Hosmer, David W., Stanley Lemeshow, and Susanne May. 2011. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. 2nd Edition. Wiley.

Hougaard, P. 1999. "Multi-State Models: A Review." *Lifetime Data Analysis* 5: 239–64. https://doi.org/10.1023/a:1009672031531.

Johnston, Marjorie C, Michael Crilly, Corri Black, Gordon J Prescott, and Stewart W Mercer. 2019. "Defining and Measuring Multimorbidity: A Systematic Review of Systematic Reviews." *European Journal of Public Health* 29 (1): 182–89. https://doi.org/10.1093/eurpub/cky098.

Lagerros, Ylva Trolle, Rino Bellocco, Hans-Olov Adami, and Olof Nyrén. 2009. "Measures of Physical Activity and Their Correlates: The Swedish National March Cohort." *European Journal of Epidemiology* 24 (4): 161–69. https://doi.org/10.1007/s10654-009-9327-x.

Lassale, Camille, Marc J. Gunter, Dora Romaguera, Linda M. Peelen, Yvonne T. Van der Schouw, Joline W. J. Beulens, Heinz Freisling, et al. 2016. "Diet Quality Scores and Prediction of All-Cause, Cardiovascular and Cancer Mortality in a Pan-European Cohort Study." Edited by Chung-Jung Chiu. *PLOS ONE* 11 (7): e0159025. https://doi.org/10.1371/journal.pone.0159025.

Le Reste, Jean Yves, Patrice Nabbe, Benedicte Manceau, Charilaos Lygidakis, Christa Doerr, Heidrun Lingner, Slawomir Czachowski, et al. 2013. "The European General Practice Research Network Presents a Comprehensive Definition of Multimorbidity in Family Medicine and Long Term Care, Following a Systematic Review of Relevant Literature." *Journal of the American Medical Directors Association* 14 (5): 319–25. https://doi.org/10.1016/j.jamda.2013.01.001.

Lee, Tae Wha, Jane Chung, Kijun Song, and Eunkyung Kim. 2022. "Incidence and Predictors of Multimorbidity among Older Adults in Korea: A 10-Year Cohort Study." *BMC Geriatrics* 22 (1): 565. https://doi.org/10.1186/s12877-022-03250-w.

Li, Xiao, Le Cai, Wen-Long Cui, Xu-Ming Wang, Hui-Fang Li, Jian-Hui He, and Allison Rabkin Golden. 2020. "Association of Socioeconomic and Lifestyle Factors with Chronic Non-Communicable Diseases and Multimorbidity among the Elderly in Rural Southwest China." *Journal of Public Health (Oxford, England)* 42 (2): 239–46. https://doi.org/10.1093/pubmed/fdz020.

Ludvigsson, Jonas F., Catarina Almqvist, Anna-Karin Edstedt Bonamy, Rickard Ljung, Karl Michaëlsson, Martin Neovius, Olof Stephansson, and Weimin Ye. 2016. "Registers of the Swedish Total Population and Their Use in Medical Research." *European Journal of Epidemiology* 31 (2): 125–36. https://doi.org/10.1007/s10654-016-0117-y.

Ludvigsson, Jonas F, Eva Andersson, Anders Ekbom, Maria Feychting, Jeong-Lim Kim, Christina Reuterwall, Mona Heurgren, and Petra Otterblad Olausson. 2011. "External Review and Validation of the Swedish National Inpatient Register." *BMC Public Health* 11 (1): 450. https://doi.org/10.1186/1471-2458-11-450.

Ludvigsson, Jonas F., Petra Otterblad-Olausson, Birgitta U. Pettersson, and Anders Ekbom. 2009. "The Swedish Personal Identity Number: Possibilities and Pitfalls in Healthcare and Medical Research." *European Journal of Epidemiology* 24 (11): 659–67. https://doi.org/10.1007/s10654-009-9350-y.

Lugtenberg, Marjolein, Jako S. Burgers, Carolyn Clancy, Gert P. Westert, and Eric C. Schneider. 2011. "Current Guidelines Have Limited Applicability to Patients with Comorbid Conditions: A Systematic Analysis of Evidence-Based Guidelines." Edited by Peter McCulloch. *PLoS ONE* 6 (10): e25987. https://doi.org/10.1371/journal.pone.0025987.

Marengoni, Alessandra, Sara Angleman, René Melis, Francesca Mangialasche, Anita Karp, Annika Garmen, Bettina Meinow, and Laura Fratiglioni. 2011. "Aging with Multimorbidity: A Systematic Review of the Literature." *Ageing Research Reviews* 10 (4): 430–39. https://doi.org/10.1016/j.arr.2011.03.003.

Marengoni, Alessandra, Albert Roso-Llorach, Davide L. Vetrano, Sergio Fernández-Bertolín, Marina Guisado-Clavero, Concepción Violán, and Amaia Calderón-Larrañaga. 2020. "Patterns of Multimorbidity in a

Population-Based Cohort of Older People: Sociodemographic, Lifestyle, Clinical, and Functional Differences." *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 75 (4): 798–805. https://doi.org/10.1093/gerona/glz137.

Marmot, M. G., M. J. Shipley, G. Rose, and B. Thomas. 1981. "Alcohol and Mortality: A U-Shaped Curve." *The Lancet* 317 (8220): 580–83. https://doi.org/10.1016/S0140-6736(81)92032-8.

McKenzie, Fiona, Carine Biessy, Pietro Ferrari, Heinz Freisling, Sabina Rinaldi, Veronique Chajès, Christina C. Dahm, et al. 2016. "Healthy Lifestyle and Risk of Cancer in the European Prospective Investigation Into Cancer and Nutrition Cohort Study." *Medicine* 95 (16): e2850. https://doi.org/10.1097/MD.0000000000002850.

McKenzie, Fiona, Pietro Ferrari, Heinz Freisling, Veronique Chajès, Sabina Rinaldi, Jordi de Batlle, Christina C. Dahm, et al. 2015. "Healthy Lifestyle and Risk of Breast Cancer among Postmenopausal Women in the European Prospective Investigation into Cancer and Nutrition Cohort Study." *International Journal of Cancer* 136 (11): 2640–48. https://doi.org/10.1002/ijc.29315.

Meira-Machado, Luís, Jacobo de Uña-Álvarez, Carmen Cadarso-Suárez, and Per K Andersen. 2009. "Multi-State Models for the Analysis of Time-to-Event Data." *Statistical Methods in Medical Research* 18 (2): 195–222. https://doi.org/10.1177/0962280208092301.

Michels, Karin B., and Walter C. Willett. 2009. "Self-Administered Semiquantitative Food Frequency Questionnaires: Patterns, Predictors, and Interpretation of Omitted Items." *Epidemiology* 20 (2): 295–301. https://doi.org/10.1097/EDE.0b013e3181931515.

Mounce, Luke T. A., John L. Campbell, William E. Henley, Maria C. Tejerina Arreal, Ian Porter, and Jose M. Valderas. 2018. "Predicting Incident Multimorbidity." *The Annals of Family Medicine* 16 (4): 322–29. https://doi.org/10.1370/afm.2271.

Naudin, Sabine, Vivian Viallon, Dana Hashim, Heinz Freisling, Mazda Jenab, Elisabete Weiderpass, Flavie Perrier, et al. 2019. "Healthy Lifestyle and the Risk of Pancreatic Cancer in the EPIC Study." *European Journal of Epidemiology*, September. https://doi.org/10.1007/s10654-019-00559-6.

Pizzi, C., B. De Stavola, F. Merletti, R. Bellocco, I. dos Santos Silva, N. Pearce, and L. Richiardi. 2011. "Sample Selection and Validity of Exposure-Disease Association Estimates in Cohort Studies." *Journal of Epidemiology & Community Health* 65 (5): 407–11. https://doi.org/10.1136/jech.2009.107185.

Report of the Joint WHO/FAO expert consultation, ed. 2003. *Diet, Nutrition, and the Prevention of Chronic Diseases: Report of a WHO-FAO Expert Consultation*. WHO Technical Report Series 916. Geneva: World Health Organization.

Singer, Merrill. 1996. "A Dose of Drugs, a Touch of Violence, a Case of AIDS: Conceptualizing the SAVA Syndemic." *Free Inquiry in Creative Sociology* 24: 99–110.

Singer, Merrill, Nicola Bulled, Bayla Ostrach, and Emily Mendenhall. 2017. "Syndemics and the Biosocial Conception of Health." *The Lancet* 389 (10072): 941–50. https://doi.org/10.1016/S0140-6736(17)30003-X.

Trolle Lagerros, Ylva, Essi Hantikainen, Daniela Mariosa, Weimin Ye, Hans-Olov Adami, Alessandra Grotta, Francesca Ghilotti, and Rino Bellocco. 2016. "Cohort Profile: The Swedish National March Cohort." *International Journal of Epidemiology*, September, dyw193. https://doi.org/10.1093/ije/dyw193.

United Nations, Department of Economic and Social Affairs, Population Division. 2020. *World Mortality 2019*. ST/ESA/SER.A/437.

Vos, Theo, Stephen S Lim, Cristiana Abbafati, Kaja M Abbas, Mohammad Abbasi, Mitra Abbasifard, Mohsen Abbasi-Kangevari, et al. 2020. "Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019." *The Lancet* 396 (10258): 1204–22. https://doi.org/10.1016/S0140-6736(20)30925-9.

Wettermark, Björn, Niklas Hammar, Carl Michael Fored, C. MichaelFored, Andrejs Leimanis, Petra Otterblad Olausson, Ulf Bergman, et al. 2007. “The New Swedish Prescribed Drug Register-Opportunities for Pharmacoepidemiological Research and Experience from the First Six Months.” *Pharmacoepidemiology and Drug Safety* 16 (7): 726–35. https://doi.org/10.1002/pds.1294.

WHO Guidelines on Physical Activity and Sedentary Behaviour. 2020. Geneva: World Health Organization.

Wikström, Katja, Jaana Lindström, Kennet Harald, Markku Peltonen, and Tiina Laatikainen. 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 (3): 211–16. https://doi.org/10.1016/j.ejim.2015.02.012.

Willadsen, Tora Grauers, Anna Bebe, Rasmus Køster-Rasmussen, Dorte Ejg Jarbøl, Ann Dorrit Guassora, Frans Boch Waldorff, Susanne Reventlow, and Niels de Fine Olivarius. 2016. “The Role of Diseases, Risk Factors and Symptoms in the Definition of Multimorbidity – a Systematic Review.” *Scandinavian Journal of Primary Health Care* 34 (2): 112–21. https://doi.org/10.3109/02813432.2016.1153242.

World Health Organization. 2002. *Innovative Care for Chronic Conditions: Building Blocks for Action: Global Report*. Geneva: World Health Organization.

Yin, Weiyao, Marie Löf, Nancy L. Pedersen, Sven Sandin, and Fang Fang. 2021. “Mediterranean Dietary Pattern at Middle Age and Risk of Parkinson’s Disease: A Swedish Cohort Study.” *Movement Disorders* 36 (1): 255–60. https://doi.org/10.1002/mds.28314.