HEALTH BEHAVIORS, COMORBIDITIES, 
AND LIFE EXPECTANCY IN MIDDLE- 
AGED MEN: THE KIHD STUDY

Mounir Ould Setti
Master’s thesis
Public Health
University of Eastern Finland
Faculty of Health Sciences
School of Medicine
May 2019
Estimating the risk of unhealthy behaviors, such as tobacco smoking, physical inactivity, alcohol drinking, and unbalanced diet, is of an utmost importance to healthcare, policy making, and health promotion. The aim of this study is to evaluate the combined effects of the main health behaviors and to develop a predictive model that permits the expression of these effects through estimating life-expectancy. The study is based on a prospective cohort of n=2682 middle-aged male participants from the region of Kuopio, Finland. Smoking, alcohol drinking, physical activity, and diet - as indicated by the Baltic Sea Dietary score - were assessed for their effect on time to all-cause mortality. After a mean follow-up of 23.3 years, smokers were associated with nearly double the risk of mortality as non-smokers (HR=1.91 95% CI 1.71 – 2.13). Mid to high quality diet (BSDS > 10) was found associated with a mortality risk reduction of up to 43% in comparison to very low-quality diet (BSDS<5) (HR = 0.57, p-value < 0.001 for BSDS from 10 to 15). Alcohol consumption (units of 100 grams per week) was associated with lower survival with a (HR=1.12 95% CI 1.08 – 1.15). In general, up to 20 years of life are to be gained by adopting an optimal healthy lifestyle from midlife on.
ACKNOWLEDGEMENTS

The more one knows the more they realize how much they do not know, for the only certainty in science is probably uncertainty.

Nevertheless, the process of gaining knowledge provides joy and illumination: feelings that I believe I was fortunate enough to experience while writing this thesis.

I thank everyone who was involved in the process.
ABBREVIATIONS

AACCII  Age-Adjusted Charlson Comorbidity Index
BMI     Body mass index
BSDS    Baltic Sea Diet Score
CAD     Coronary Artery Disease
CCI     Charlson Comorbidity Index
CI      Confidence Interval
CIRS    Cumulative Illness Rating Scale
COPD    Chronic obstructive pulmonary disease
CPS     Comorbidity-polypharmacy score
CVD     Cardiovascular Disease
DALY    Disability-adjusted life year
EROS    European Registers of Stroke
FHS     Framingham Heart Study
GBD     Global Burden of Disease Study
GOLD    Global Initiative for Chronic Obstructive Lung Disease
GUI     Graphical user interface
HDI     Human Development Index
HHD     Hypertensive heart disease
HLE     Healthy Life Expectancy
HR      Hazard Ratio
IARC    International Agency for Research on Cancer
ICD     International Classification of Diseases
IHME    Institute for Health Metrics and Evaluation
MET     Metabolic equivalent
MI      Myocardial infarction
NCD     Noncommunicable disease
NIAAA   National Institute on Alcohol Abuse and Alcoholism
OECD    Organisation for Economic Co-operation and Development
PH      Proportional Hazards
QALY    Quality-adjusted life year
SD      Standard Deviation
SDI     Socio-demographic Index
SIDS    Sudden infant death syndrome
TIA     Transient ischemic attach
WHO     World Health Organization
1 INTRODUCTION

Historically, when the focus of public health was on infectious diseases which were leading mortality rates, death from cardiovascular diseases was regarded as a natural consequence of aging and a logical limit of life-expectancy. Before cardiovascular risk and the term risk factor became a thing, cardiovascular disease was considered an out of reach domain and an unpreventable phenomenon that healthcare can merely observe, follow-up, and, if possible, palliate. The pathogenesis and advancement of such disease was simply regarded as idiopathic. Up until recent times, some physicians still reported ‘old age’ on death certificates as a cause of death from cardiovascular disease, and probably some other non-communicable diseases (Oregon Health Authority 2013). It was not until results from epidemiological studies on the association between health-related behaviors and cardiovascular diseases came out that the scientific community learned that control measures can be adopted against mortality from cardiovascular diseases.

The main goal of public health as a science is to first find, through evidence, what is causing harm to people’s health, and then to try to prevent this harm from happening. The ultimate result would be the improvement of the population’s quality of life and life expectancy. Preventing diseases is the fruit of the combined efforts of a multitude of stakeholders among which health promotion plays a major role (World Health Organization 2012).

Health promotion is the process that aims to improve public health by helping people gain control over their health and the factors affecting it (World Health Organization 2005). While it is meant as a concept to achieve equity in health by working on the determinants of health in all their dimensions, one of its main areas of action is people’s behavior.

This study is meant to identify the major behaviors affecting health and life expectancy as an outcome and to rank them in terms of importance as to help health promotion informatively set its priorities.
2 LITERATURE REVIEW

2.1 Life expectancy

Life expectancy is defined as the length, in years, on average, a person is expected to live provided mortality rate does not change. The most commonly used life expectancy measure is life expectancy at birth which refers to life expectancy of a newborn. As an indicator, it reflects the overall mortality level of a population and it is usually estimated based on demographic statistics of the previous generations.

Historically, life expectancy varied significantly throughout time and populations from about 20 years in the Neolithic Age and 26 in the Bronze and Iron Ages (Galor & Moav 2005, 2007), to about 69 years in Renaissance Italian philosophers (Benet 1972) and an average of 84 years in 11th century Muslim scholars (Bulliet 1983). However, it is important to draw attention to the contribution of violent deaths and infant mortality in those numbers (Rowbotham & Clayton 2008). For instance, excluding those who died violently, and controlling for infant mortality, by for example considering life expectancy at the age of 5 or 15 years instead of at birth, reduces the variation of life expectancy throughout the past three thousand years in a remarkable manner. Montagu found that it was common, during the period from 650 BC to 100 BC, to live for about 72 years (Montagu 1994) which is not far from today’s expectancies. Nevertheless, these numbers only applied to ‘men of achievement and fame’, as Montagu noted, and do not represent the general population. JP Griffin, in his letter to the Journal of Royal Society of Medicine, notes that the situation was different on the side of women whom life expectancies have been subject to a wide range of variation over the past six hundred years (Griffin 2008). Data from the British Ducal Families show that women lived on average 48 years in the 16th and 17th century and 57 years in the 18th century (Hollingsworth 1957).

While it is a common belief that medical advancements are the reason, public health measures and policies had the biggest contribution in the increase of life expectancy since the 1800s. The period from the late decades of the 19th century to the early decades of the 20th, in which infectious diseases such as tuberculosis, polio, diptheria, and smallpox were major causes of mortality, was marked by what is called the “First Public Health Revolution”: a struggle against infectious disease before the age of antibiotics. This fight was won with public health measures that eliminated a multitude
of causes of death through environmental actions including proper management of water, sewage, garbage, and sanitation, and through social policies such as those related to child labor and improvement of literacy and nutrition. And then thanks to vaccination, that most common cause of infant mortality and childhood death at that time, became a rare cause of mortality today (Health 1988). Measuring the impact of public health interventions was only possible because of the availability of statistics on rates and specific causes of death (Bunker et al. 1994, Novick 2008).

In fact, making informed judgements regarding the performance of governmental efforts to improve public health and health care depends, almost solely, on the availability of statistical indicators. The need to have health-outcomes focused indicators is an important element of efficiency and effectiveness of public health measures. Omitting that and focusing on determinants and policies would widen the gap between the efforts put in improving healthcare and the population’s status of health. In that sense, life expectancy constitutes a very targeted and simple, or probably the simplest, indicator of health status to evaluate inequalities and guide resource allocation. Life expectancy is also often used, in combination with other factors, to create other scores and indicators. The Human Development Index (HDI), for example, combines a life-expectancy index with an education index and a national growth index allowing comparison between countries’ development in a way centered on individuals. HDI permits to reveal contrasts on which policies work best for the people (Stiefel et al. 2010, United Nations Development Programme 2016).

2.1.1 Quality of life

While change in life expectancy is considered a good ‘guess’ of changes in health status, quality of life, constituting an entangled component with health status, is just as important to measure as life expectancy. As a step forward from the status of being alive, living a healthy, productive and enjoyable life, defines the concept of quality of life which is understood roughly the same as what is meant by the term “health status”, and the term “well-being” although there is no consensus to give a single definition for this last and it is usually more related to mental health. Different measures try to assess, in a quantifiable fashion, quality of life. For instance, some of its most popular measures are probably Quality-adjusted life years (QALYs) which was introduced in the 70s and Disability-adjusted life years (DALYs) which came two decades later. The former is a
measure of life expectancy combined with quality of life. It is calculated by multiplying each of
the years of life expectancy by a coefficient of quality of life, of that year. Each year is called a
QALY. This coefficient, or weight, goes from 0: dead, to 1: in perfect health, and thus, QALY is
based on a maximization criterion and walk hand in hand with the utilitarian philosophy.
Consequently, and as QALY gained in popularity and research interest, it became the gold-standard
in analyzing cost-effectiveness and many other (Drummond et al. 2015). DALYs on the other hand,
derived from the framework of QALYs, take these lasts to a deeper level accounting for the age at
which the disability is happening (Sassi 2006). Nevertheless, since 2010, for its Global Burden of
Disease Study (GBD), the World Health Organization (WHO) has adopted a simpler form of
estimation of DALY that does not account for age (World Health Organization 2013).

Expressed as a percentage of overall life expectancy, Healthy Life Expectancy (HLE) is another
measure that combines, in its assessment of health status, life expectancy with quality of life. It
seems to be a more intuitive and meaningful stand-alone measure of health. It accounts for age in
its weighing allowing comparison of populations with different age distributions and also to give
an image as of the development of the compression and expansion of a population’s morbidity
(Stiefel et al. 2010).

2.1.2 Compression of morbidity

In order to portray the concept of compression of morbidity as introduced by James F. Fries in a
remarkable paper published in 1980, “Think about two points on a typical human lifespan, with the
first point representing the time at which a person becomes chronically ill or disabled and the
second point representing the time at which that person dies. Today, the time between those two
points is about 20 years or so. During the early portion of those years, chronic disease or disability
is minor, but increases nearer to the end of life. The idea behind compression of morbidity is to
squeeze or compress the time horizon between the onset of chronic illness or disability and the time
in which a person dies.” (Fries 1980). The logic of the hypothesis is that all non-traumatic deaths
are due to an illness, that most illness is chronic, that aging raises the probability of developing
chronic illnesses, and thus, chronic illnesses happen in later life. The theory suggests that delaying
the onset of chronic illnesses might lighten their lifetime burden provided that this delay is bigger
than the increase of life expectancy. In other words, compression of morbidity happens when
disability is reduced at a higher rate than the reduction in mortality. It is probably also important to note that the decline of disability is not necessarily synonym of a lower incidence of chronic illnesses.

Compression of morbidity can also be considered as one of the three possible scenarios of the becoming of the present lifetime morbidity as presented by Fries in a more recent article (Fries 2003) and illustrated below (Figure 1).

![Figure 1. Possible scenarios for future morbidity and longevity, inspired from Fries (2003)](image)

It was a common belief a few decades ago that the upcoming increase in life expectancy would lead to the unfortunate expansion of morbidity or Life Extension scenario. This “failure of success” (Gruenberg 1977) situation is based on the fear that the extra time gained in the future due to the development of medicine is mostly going to be spent in a miserable state of chronic illness and that the growing number of older adults is only a synonym of a heavier burden. In that sense, the concept of Compression of Morbidity was seen as the ideal scenario, probably even a ‘too good to be true’ kind of scenario and a call to underestimate the necessary amount of preparedness to face the future.

However, Fries, based on data from US national surveys and observational longitudinal studies (Singer & Manton 1998), demonstrates that:
1- The trend in the rise of life expectancy at 65 years of age, since 1980, is getting more and more slower and it is projected to continue in that sense for a while.

2- For multiple reasons such as the improvement of medical care and the reduction of unhealthy behaviors, the slope of the decline in morbidity is getting steeper.

3- The rate of decline of disability is greater than the rate of decline in mortality.

These conditions indicate a rapid taking place of the scenario of compression of morbidity (Fries 2003).

Similarly, Hubert et al. studied 2328 subjects in a cohort of 13 years obtained from the College Alumni Study. They have annually collected data on health behavior, medical history and use, as well as physical disability. This last was assessed using the validated functional status measurement tool: Health Assessment Questionnaire. On the other hand, the participants’ health-related behaviors have been measured such as vigorous physical activity, smoking, and being under- or overweight. Spline regression models were then fitted through generalized estimating equations for the purpose of testing the changing rate of disability. This statistical method was mainly chosen because it requires no assumption as of the distribution of the data. Bootstrapping was then used to validate the results and to make a conclusion about the compression of morbidity. This methodology fits ‘knot points’, at three-month intervals in this case, to the regression spline and allows comparison of rate of change between before and after each of the knots. The study confirmed the compression of morbidity associated to low-risk health behaviors and concluded that the consequent life-span increase of a healthy lifestyle comes with an even greater delay of disability (Hubert et al. 2002).

In short, the compression of morbidity hypothesis, which some prefer to call -for better accuracy- ‘compression of disability’, assumes that the same forces that postpone death would cause an even greater postponement of the onset of disability. However, the concept did not go uncriticized since Fries proposed it almost four decades ago. Recent studies on middle age people have shown conflicting results as of the health trends of the next generations of older adults. A report published by the Center for Retirement Research at Boston College found that healthy life expectancy might not be so different in the future (Munnell et al. 2008) realizing either the previously presented scenario of Life Extension, or the Shift to the Right: a scenario that is similar to the situation of
'dynamic equilibrium’ proposed by Manton who presumed that chronic diseases will get milder in severity but longer in duration in a trend of increasing life expectancy keeping constant the proportion of healthy life years (Manton 1982). Soldo et al. were even more skeptical about the future trends of end-life well-being (Soldo et al. 2006).

The fact is, the occurrence of morbidity compression has not been constant; at least, not everywhere. A study evaluating trends of quality of life and disability in the older adults of 12 OECD countries found that disability is diminishing in only few of the studied countries and that includes Finland and the United States on which Fries based his studies. While a few other countries report an equilibrium between mortality and disability, Japan and Sweden have been found to have a trend that is more toward the increase of disability (Lafortune & Balestat 2007). However, these studies are too broad to deny the hypothesis that the same process, also called healthy aging, that delays mortality would improve quality of life. Due to the complexity of societies and the imbrication of an uncountable number of factors -notably socioeconomic-, making inference on a biological process using demographical studies might have a lot of drawbacks regarding its plausibility. Recent studies of the biology of aging at cellular and molecular levels have found that postponement of senescence is a very possible and realistic goal (Kirkwood & Austad 2000, Sierra et al. 2009) and that delaying the aging process will both delay mortality and ameliorate quality of life (Goldman et al. 2013).

Based on that, it is relatively safe for us to assume that lifestyle-related increase of life-span could be accompanied by a compression of morbidity and a better quality of life, and thus, focus our study on assessing the effects of lifestyle and health-related behaviors on life expectancy. Especially that, in cancer for example, functional status have not been found to correlate with the stage of cancer, or with comorbidity (Extermann et al. 1998). However, future research could also evaluate the effects of lifestyle and behavior on disability and quality of life and verify the occurrence or not of compression of morbidity.

2.2 Comorbidities

Non-communicable diseases (NCDs), in contrast to infectious diseases, are defined by their non-transmissible nature. Although some of them might be rapidly lethal, NCDs refer to diseases that usually develop slowly over a long period of time requiring chronic care management. In addition
to their detrimental effect on quality of life, these chronic diseases, also designated sometimes in
clinical practice as comorbidities especially referring to the chronic conditions of a patient
presenting with an acute event, account for about 30% of all deaths globally (World Health
Organization et al. 2011), and they do significantly affect mortality rates and longevity, especially
if combined.

Multimorbidity, defined as the presence of two or more chronic diseases simultaneously in a
person, can dramatically affect life expectancy estimates. Authors from the Emerging Risk Factors
Collaboration (2015), based on a huge study that included 91 cohorts, observed a reduction of 15
years of life expectancy in patients with cardiometabolic multimorbidity at the age of 60, and an
even greater reduction of life expectancy of 23 years in patients with cardiometabolic
multimorbidity at the age of 40.

Although this fact is well established, attention to the growing impact of the combination of
multiple chronic conditions has not been given enough justice until, probably, recently. For
instance, in the United States, experts estimated in 2000 that a quarter of the US population would
suffer of multimorbidity by 2030, however, the prevalence of multimorbidity there already reached
28 percent in 2006 (DuGoff et al. 2014). The situation is not very different in Europe, 50 million
person live with multiple chronic diseases and the numbers are projected toward the rise (Rijken et
al. 2013).

The term comorbidity, introduced in 1970 (Feinstein 1970), refers to the presence of a disease
independently of the main disease under study differing from the term multimorbidity in the sense
that this last is wider in its view of health status and does not necessarily attach a health problem
to a studied disease. This implies that comorbidities are taken more as prognosis determinants for
an index disease while in multimorbidity the interest is on the effect of their co-occurrence on the
individual (Batstra et al. 2002). In our study, our interest is the effects of the main health-related
behaviors on life-expectancy. Therefore, we are not interested in an index disease but on any
disease that will end up causing mortality. Thus, we will term a comorbidity, any disease with
which the patient is presented at baseline.
Not many studies examined how multimorbidity affects life expectancy, and the ones that did, focused on a set of specific illnesses. However, it has been established that multimorbidity is associated with premature mortality, disability, and worse quality of life, especially in vulnerable population groups in which multimorbidity is even more prevalent (Fortin et al. 2007). Cho and colleagues (2013), for example, studied the life expectancies of more than 400 thousand randomly selected beneficiaries of Medicare who were aged 66 years or older between 1992 and 2005 excluding those with a history of cancer. They have found that people with comorbidities have a considerably lower life expectancy than the general population of the same age. DuGoff et al. (2014) also studied a sample from Medicare (N = 1,372,272) aged 67 and older and observed significant differences in life expectancy between individuals with no chronic conditions and individuals with multiple chronic conditions at the same chronological age. It is therefore relevant to our study to take comorbidities and their combined effect into account for a better assessment of the consequences of health-behaviors on longevity especially that both the nature of comorbidities and their severity represent important confounders and may affect survival studies (Porta 1997).

What we are going to present next is a group of few non-communicable diseases that happen to be physiopathologically inter-related and that account for most of the disease burden and 73% of mortality in Europe. More than three quarters of the lost DALYs in Europe would be attributed to the consequences of NCDs (World Health Organization 2006).

2.2.1 Cardiovascular disease

Cardiovascular disease (CVD) refers to the diseases and conditions that touch the heart and the blood vessels, notably those of the brain (World Health Organization et al. 2011). CVD is the leading cause of death worldwide and, globally, one of the heaviest public health concerns. According to WHO, an annual 17.9 million deaths (2017) is attributed to cardiovascular diseases accounting for a third of all deaths worldwide. This rate is trending toward the raise and an annual mortality of 23.6 million is projected by the year 2030 (World Health Organization et al. 2011).

One of the first follow-up studies that were set with the intention to properly study the epidemiology of cardiovascular disease in a large population was probably the attempt of Sir James Mackenzie in the 1920s to follow-up the health of the entire population of a town in Scotland. The study was intended as a long-term study but it was not carried out to a conclusive stage due to the
retirement of Mackenzie (Mackenzie & Orr 1926). There was no other study on the subject of cardiovascular disease epidemiology until probably 1947 when the Minnesota Business and Professional men study (Keys et al. 1963) started following-up, for fifteen years, the health of 300 middle-aged men. Prior to that, understanding of cardiovascular disease was so limited that Franklin Roosevelt, president of the USA from 1933 to 1945, during his presidential campaign in 1932, was reported having a ‘normal’ blood pressure of 140/100 mm Hg. A year later, the president was then appointed an ear, nose, and throat specialist to be his main physician because a headache from which he was suffering was thought to be his main health problem. At the age of 59, the President’s blood pressure rose to 188/105 mm Hg and was still considered normal “for a man of his age”. It was not until three years later that, after that the President’s daughter saw that his health is deteriorating and insisted on a second medical opinion, that the President’s first diagnosis of hypertension and cardiac failure was given. The president had died a year later (Bruenn 1970, Bumgarner & Floyd 2004) sharing the fate of half of the Americans at that time in their death from CVD disease (Kannel 1990).

What is considered the most iconic and extensive long-term epidemiologic study of coronary artery disease, up until today, is probably the Framingham Heart Study (FHS) which began in 1948. The Original Cohort followed-up 5209 respondents randomly selected from the slightly bigger population of the aged 29 to 62 adults of the town of Framingham (Dawber et al. 1951). FHS then included other cohorts that studied the offspring of the original participants and then a cohort of the third generation and other cohorts. The original cohort kept running until the 32nd exam in April 2014. Findings from FHS contributed, through identifying those at most probable risk of having a future cardiovascular event, to a shift in focus of public health from treatment to prevention of cardiovascular disease (Mahmood et al. 2014).

CVD mainly comprises (World Health Organization 2017):

1- Coronary Artery Disease (CAD)
Also known as coronary heart disease and ischemic heart disease, CAD represents the most prevalent form of cardiovascular disease. In 2013, CAD was the most common cause of death worldwide (Feigin et al. 2017). It mainly includes angina and myocardial infarction which are
typically the consequence of an ischemia of the coronary arteries supplying the heart muscles through the process of atherosclerosis (Wong 2014).

Although this last has widely been the subject of biology education at different levels, understanding of the science underlying its mechanism has changed so dramatically over the past 20 years (Hopkins 2013) that we estimate a brief update on the recent insights of the physiopathology of atherosclerosis would be highly relevant to the scope of this literature review.

Atherogenesis, or the process of constitution of atherosclerosis, refers to the abnormal accumulation of fatty plaques, cholesterol, and other debris in the intima of an artery. This build-up of material constitutes the atheroma, or the atheromatous plaque, and causes the wall of the artery to swell and its caliber to narrow which may reduce the flow of blood through it (Hopkins 2013). Atherosclerosis develops in a chronic and progressive fashion through a mechanism of inflammatory nature. Based on an early paper of Hopkins and Williams (1981), when it manifests clinically, atherosclerosis would have passed through four main phases: First, the inflammatory process is initiated by a recurrent damage to the endothelium of the artery. This inflammation of the endothelium then, through a mechanism involving platelets, promotes the deposition and retention of lipoproteins, and the build-up of smooth muscle cells in the intima underlying the inflamed endothelium. As a third phase, the constituted plaques get remodeled and the disease advances through fibrosis and thrombosis with an enlargement of the necrosis area. The final phase is represented by the abrupt occurrence of clinical events as a consequence of major obstruction of the concerned artery which is generally triggered by rupture of the atheromatous plaque and thrombosis.

However, these clinical events, typically myocardial infarction (MI), unstable angina, cardiac arrest, or ventricular fibrillation, might also occur as a result of an acute coronary syndrome after a strenuous activity on a physically untrained heart (Mittleman et al. 1993), or an emotionally stressful or exciting situation on an electrically unstable myocardium (Figueredo 2009). An example of the former situation would be firefighters who are 12 to 136 times more at risk of a heart attack when suppressing fire than when performing nonemergency duties (Kales et al. 2007). For the latter situation, cardiovascular events occurring in patients watching what they consider as important football matches is used as a typical example of emotional stress triggers heart attacks.
On a study by Wilbert-Lampen et al. (2008) in which 4279 patients were assessed for acute clinical events, among the previously listed, occurring in the Munich area during the World Cup 2006 and during a few control periods. The team of researchers found that watching stressful football matches might triple the risk of acute cardiovascular events in men and double it in women especially if there is a known history of CAD. However, the same team reassessed the same data for the number of deaths due to myocardial infarction and did not find a statistically significant increase of MI-attributed mortality (Wilbert-Lampen et al. 2011). Another study reexamined incidence of cardiac events within the same population and during the same World Cup but using an in-depth analysis of a larger and more representative sample. The Austrian team found no association between emotional stress from watching Football matches and cardiovascular events (Niederseer et al. 2013). Nevertheless, a growing body of research supports the evidence that emotional stress contributes in the rise of cardiovascular events, especially in situations such as earthquakes (Leor et al. 1996), or war (Bergovec et al. 1992, Chi et al. 2003). In addition, loss of a close relative or a significant person in one’s life has been found to multiply the risk of incidence of acute MI by 21.1 on the first day following the loss (Mostofsky et al. 2012). Emotional stress is, thus, considered as a precipitating factor of a clinical event happening on an already advanced CAD not fully acting as an etiology but more as what screening is to lead-time bias.

A number of risk factors of CAD are well determined nowadays. Although genetics play a big role in the pathogenesis of the disease and in the predisposition to a certain type of risky health behaviors (McPherson Ruth & Tybjaerg-Hansen Anne 2016), a good proportion of the burden of CAD is linked to modifiable risk factors such as lack of exercise, smoking, alcohol, and poor diet. These risk factors model other risk factors, said metabolic risk factors, which themselves account heavily in the pathogenesis of the disease, such as hypertension, obesity, high blood cholesterol and high blood triglycerides. These same risk factors play a role in other cardiovascular diseases and non-communicable diseases (World Health Organization et al. 2011) adding more legitimacy to the global focus of public health authorities on improving health-related behavior.

2- Cerebrovascular Disease

In 2013, cerebrovascular disease was the form of cardiovascular disease that follows coronary artery disease in term of mortality with a percentage of 35% of the total mortality attributed to
cardiovascular disease. Cerebrovascular disease was also the second most common cause of death and the third responsible of disability in the world (Roth et al. 2015).

Cerebrovascular disease refers to different conditions consequent to an affection of the blood vessels that are supplying the brain and manifest mainly with a stroke. Strokes, or cerebrovascular accidents, second leading cause of death worldwide with an annual mortality surpassing 6 million (World Health Organization et al. 2011) are mainly of three types: ischemic strokes, hemorrhagic strokes, and transient ischemic attacks (TIAs or mini strokes).

Classically, stroke has a clinical definition. It is defined by its focal (and sometimes global) neurological manifestations that have a vascular etiology. The clinical manifestations tend to differ depending on the localization of the lesion defining a variety of syndromes (Sacco et al. 2013). Ischemic strokes and TIAs, representing more than 80% of stroke cases, are both caused by an obstruction type of disruption of cerebral blood flow. The latter is defined by the brevity of its manifestations and the transience of its damage (American Heart Association 2018b), and thus, for a stroke to qualify as a TIA, all neurological symptoms need to resolve within 24 hours and no brain damage shall be demonstrated. Ischemic strokes, on the other hand, result in brain infarctions and permanent damage (Smith et al. 2016). Both ischemic and mini strokes typically have atherosclerosis as an underlying mechanism. The cerebrovascular event could either be due to a locally formed thrombus over a local atherosclerotic plaque and defining a thrombotic stroke, or due to a traveling embolus originating distantly (typically from the auricles of the heart) and defining an embolic stroke (American Heart Association 2018b). Hemorrhagic strokes on the other hand, although only accounting for 13% of stroke cases, are responsible of over 40% of cerebrovascular disease mortality. Hemorrhagic strokes present in two main types: intracerebral hemorrhage and subarachnoid hemorrhage depending on whether the bleeding is within the brain tissue (i.e intraparenchymal or intraventricular) or within the cranial volume but outside the brain tissue (American Heart Association 2018a, Roth et al. 2015).

Epidemiologically, the distribution of cerebrovascular disease is function of many factors. Sex differences in cerebrovascular disease incidence are evident. While women have more incidence of strokes in absolute numbers, since these last tend to have a higher life-expectancy, when adjusting for age, men tend to have 50% excess risk of stroke compared to women according to
studies on the European Registers of Stroke (EROS) (Heuschmann et al. 2009). Geographically speaking, although not many proper studies have been conducted in Africa, low- and middle-income countries tend to have a clearly higher incidence of cerebrovascular disease in comparison to high-income countries (Seshadri & Debette 2016). Geographical variations are also noted in term of the distribution of subtypes of cerebrovascular disease. Incidence rates of hemorrhagic strokes in low- and middle-income countries, for example, have been found to be double those of the high-income countries (Feigin et al. 2009). Disparities in distribution are attributed to different distribution of risk factors, notably the vascular ones, however, new data from a Chinese study suggest that the previous studies might lack in methodology to prove such conclusions (Fang et al. 2012). Mortality also, and obviously, differs by subtype of cerebrovascular disease, and hemorrhagic strokes have about 2.5 to 3 times the mortality rate of ischemic strokes. Prognosis, on the other hand, in addition to a variation by subtype, also depends on the other comorbidities accompanying the incident (Seshadri & Debette 2016).

In term of etiology, cerebrovascular diseases are modulated by several modifiable and non-modifiable risk factors. For instance, genetic predisposition to stroke has long been established and familial strokes have marked the lines of history with some cases such as that of the Elamite kings who lost their dynasty when stroke hit two of their successive kings 2700 years ago (Ashrafian 2010). Congenital diseases such as sickle cell anemia and Anderson–Fabry disease are typical congenital diseases that have been proven to have a clear effect on the risk of stroke (Meschia et al. 2011). Some other congenital conditions, such as familial amyloid angiopathies and Osler-Weber-Rendu, could also manifest as cerebral vascular malformations with high risk for hemorrhagic strokes. Risk of such malformations is also higher with connective tissue defects such as Ehler-Danlos and Marfan disease. The genetic diseases that we have mentioned are monogenic diseases, and although they are rare individually, they represent a good proportion of the incident strokes if we consider the sum of their rates (Leblanc et al. 2009). However, the forms of cerebrovascular disease that are more frequent tend to be associated, not with monogenic conditions, but more with a probabilistic “complex genetics” mechanism that rather makes the individual more vulnerable to certain environmental exposures affecting directly disease risk. Complex genetics could also make certain diseases or their complications more severe, alter the recovery or increase the susceptibility to recurrence (Seshadri & Debette 2016).
Nevertheless, the clearly avoidable risk factors for cerebrovascular disease are well-known. While medical research might provide more knowledge on the pathophysiology of the genetic risk factors, and eventually reveal targets for medical treatments, according to studies that illustrated the combined genetic and environment etiology of strokes, health promotion efforts could substantially reduce the effects of the environmental component and lower mortality (Flossmann et al. 2004). These environmental factors include diet, tobacco and alcohol consumption, physical inactivity, estrogen therapy, and also the effect of climate and the exposure to some pollutants (Seshadri & Debette 2016).

Many of these factors modulate the most important risk factor of cerebrovascular diseases: hypertension. For instance, high blood pressure, a mostly asymptomatic condition that touches more than a quarter of the world population (Ibrahim & Damasceno 2012), is thought to be responsible of nearly half of all the risk for stroke (Whisnant 1996). These factors are also associated with atherosclerosis and hypercholesterolemia, which themselves modulate the risk of developing other stroke risk factors such as atrial fibrillation and coronary artery disease (Benson & Sacco 2000). On the other hand, atherosclerosis and some vascular anomalies of the cerebral arteries are often a consequence of diabetes (Gorelick & Alter 1994), which itself is associated with many of the previously mentioned environmental risk factors of cerebrovascular disease.

Data on the burden of stroke tends to be scarce but results from the Global Burden of Disease study indicate that 50 million disability adjusted years DALYs were lost globally in 2005 because of cerebrovascular diseases. This rate, in the older adults population, represented 13% of the global burden of disease (Johnston et al. 2009, Strong et al. 2007). While the epidemiology of stroke has been rapidly changing over the past few decades, its trend of mortality and disability, in absolute numbers, is rushing toward the increase (Feigin et al. 2015). Stroke remains a universal pandemic and more efficient measures to reduce its global burden are urgently needed.

3- Other Cardiovascular diseases

CAD and Cerebrovascular Disease, when their rates are combined, share almost 80% of the total mortality from cardiovascular diseases. The pie chart below summarizes the shares of the main subtype of cardiovascular disease (World Health Organization et al. 2011).
Among the subtypes of cardiovascular disease, following CAD and cerebrovascular disease, hypertensive heart disease (HHD) would rank third in term of mortality (World Health Organization et al. 2011). HHD lacks as a condition a clear and comprehensive definition, but it mainly refers to the chronic effects of high blood pressure on the heart, and according to some recent definitions, HHD is directly linked to left ventricular hypertrophy signing its typical hemodynamic consequence. For instance, hypertension directly and independently determines left ventricular hypertrophy: a feature commonly present on some heart diseases at the beginning of their evolution toward an often fatal cardiovascular events (Kannel 1990). However, left ventricular hypertrophy is not the only consequence of HHD. In addition to the hemodynamic complications of HHD which might lead to heart failure, Gonzalez-Maqueda et al. proposed a new classification...
of HHD to also include ischemia and arrythmia since they are also modulated by hypertension and HHD. The former is a common precursor of cardiac arrest and the latter, with atrial fibrillation as its most common form, could be responsible of deadly emboli (Gonzalez-Maqueda et al. 2009). However, since hypertension is also involved in the pathogenesis of CAD, among other heart-related diseases, it is difficult to properly determine the burden of HHD.

Other cardiovascular diseases worth mentioning include: peripheral arterial disease, which is the disease of the vasculature of the limbs; rheumatic heart disease: consequence of the damage to heart valves from streptococcus infection; deep vein thrombosis, which can lead to pulmonary embolism through the dislodgement of blood clots from leg veins (World Health Organization 2017).

2.2.2 Hypertension

Hypertension, also called high blood pressure and thought that it would be responsible of nearly 10 million deaths worldwide in 2018 (Forouzanfar et al. 2017), has been called the silent killer because it tends to remain asymptomatic until it causes a critical event (Benson & Sacco 2000). As it has been previously shown, hypertension is directly connected to an increased risk of many cardiovascular diseases, notably coronary artery disease and cerebrovascular disease in which it is considered the single most important modifiable risk factor (Seshadri & Debette 2016).

The World Health Organization defines hypertension as the condition in which high pressure is continuously present in the vascular system creating resistance against the heart when it is trying to pump blood through the arteries (World Health Organization 2018a). Hypertension, as practically defined by Evans and Rose (1971), refers to the level of pressure inside the arteries “above which investigation and treatment do more good than harm”. This level of blood pressure has been changing throughout the years and according to different schools and guidelines. For instance, while a systolic blood pressure of 130 up to 139 mmHg was an accepted range of blood pressure prior to 2013, it is agreed nowadays that a threshold for treatment would be a systolic blood pressure of 130 mmHg or a diastolic blood pressure of 85 mmHg. Although new epidemiological evidence from the Global Burden of Disease Study 1990 – 2015 associates cardiovascular risk with a systolic blood pressure of 110 mmHg and above, as there is no proof that lowering systolic blood pressure under 140mmHg would have any benefit on mortality.
(Forouzanfar et al. 2017), the diagnosis of hypertension is not imposed for levels of systolic blood pressure of 129 mmHg and below and diastolic blood pressure of 84 mmHg and below. The latest guidelines recommend that only a systolic blood pressure of 140 mmHg and above or a diastolic blood pressure of 90mmHg and above would define hypertension. This clinical definition of hypertension is based on, what may be perceived as somehow arbitrary, cut-off blood pressure values, however, these values are based on evidence that treatment of patients with higher than these values will be beneficial. The definition is, thus, endorsed in clinical practice for these pragmatic reasons as to simplify diagnosis process and treatment decision (Williams et al. 2018).

Epidemiologically, despite the absence of proper data on the prevalence of hypertension in developing countries, it has been found that these lasts have about the same prevalence of hypertension as developed countries (Ibrahim & Damasceno 2012), and thus, the estimate is that nearly a sixth to more than a third of the world’s population is affected by hypertension (Poulter et al. 2015). A recent pooled analysis of 1479 studies with 19 million participants has shown that the global prevalence of hypertension in 2015 was about 24% in men and 20% in women (NCD Risk Factor Collaboration (NCD-RisC) 2017). The current tendency of hypertension distribution is that a higher prevalence is observed in urban areas in contrast to rural ones, and also in African ethnicity (Ibrahim & Damasceno 2012). However, contrasts in socio-economic status might explain these differences biasing the previous observations (Agyemang et al. 2009) and this general trend of high hypertension prevalence is consistent across all the regions of the world (Chow et al. 2013, Kearney et al. 2005).

When we refer to hypertension, we generally mean essential hypertension, also called primary hypertension. Essential hypertension is the most common form of high blood pressure and is defined by its unidentifiable cause, hence the appellation idiopathic hypertension. For instance, 5 to 10% of all cases of hypertension are classified as secondary hypertension as they have a clear etiology (Rimoldi et al. 2014). Hypertension in these cases may then be curable by managing the underlying cause. However, as it is very difficult to screen all patients with high blood pressure for secondary hypertension, and since this last tends to start earlier in life as a consequence of a renal, endocrine, or iatrogenic origin, secondary hypertension is usually suspected in patients younger than 40 years of age, patients with a hypertension that is very sever or resistant to treatment, or in the presence of suggestive symptoms (Oparil et al. 2003, Poulter et al. 2015, Williams et al. 2018).
Conversely, primary hypertension appears at more advanced ages as a consequence of some complicated processes involving environmental and genetic factors. For instance, the role of genetic factors in this disease is incontestable with an estimate of about a third of variance in blood pressure explained by genetics (Poulter et al. 2015). Classic genetics with a single gene or loci to explain a disease do not explain the physiopathology of essential hypertension. Individual loci have not found to explain more than 1 mmHg of variation in systolic blood pressure or more than 0.5 mmHg in diastolic blood pressure (Munroe Patricia B. et al. 2013). Affordable large genome-wide studies have enabled the identification of about 120 loci responsible of variation in blood pressure and risk of cardiovascular diseases. Although these discoveries are a breakthrough in our understanding of the biology of hypertension, they only explain 3.5% of the genetic component of the physiopathology of hypertension (Warren et al. 2017) making genetic testing practically useless in routine clinical care (Williams et al. 2018).

Essential hypertension’s etiopathology remains multifactorial, and in addition to the genetic component, many hypotheses also attempt to explain its genesis. Exposure to psychological stress for example is thought to contribute in essential hypertension by activating the sympathetic nervous system. This last is involved in another mechanism called vascular reactivity which implies that hypertension patients’ vasculature manifests a greater response to vasoconstrictors than other patients. These patients would thus be more vulnerable to the effect of psychological stress on hypertension. Another implicated mechanism involves the process of vasoconstriction, the aldosterone renin angiotensin system, and the metabolism of sodium. A main actor in this pathway would be angiotensin II: a peptide hormone involved in the mechanism of vasoconstriction and sodium retention. This hormone’s production is also thought, by enhancing the formation of the oxidant superoxide, to contribute in the effect of oxidative stress increasing then its harmful effects. Other mechanisms of hypertension pathogenesis involve structural remodeling and possible anatomo-histological abnormalities of the blood vessels, particularly the endothelium. This last secretes endothelin, a peptide secreted in the lumen of the vasculatures with significant peripheral vasoconstriction and vasodilation capabilities. Some hypertensive patients have increased levels of serum endothelin, notably Americans of African descent (Oparil et al. 2003).

Understanding these pathophysiological mechanisms might help in the development of more targeted antihypertensive therapies. While the existing antihypertensive medications work on these
pathways, and nuances in the pathophysiology of Essential hypertension in different categories of hypertensive patients might explain the variation of response to different medications, none of the used therapeutic guidelines base the choice of antihypertensive medication on any understanding of the pathophysiologic mechanism underlying individual patients’ hypertension. For instance, different clinical guidelines have been recommending different strategies of chemical management of hypertension. However, most of them are based on generically choosing a therapy and then experimenting with different dosages and drug combinations until blood pressure control is achieved. The mainly used antihypertensive medications belong to the following five major classes of drugs: Diuretics, Beta-blockers, Calcium antagonists, Angiotensin-converting-enzyme inhibitors, and Angiotensin II receptor blockers. While previous clinical guidelines start with a monotherapy and the current ones recommend starting directly with a combination, the initial choice is, anyway, based on a core recommended therapy with the encouragement to deviate from it as to try to benefit a comorbidity with the same antihypertensive medication if a specific indication presents. Example: privileging a beta-blocker for a patient requiring heart-rate control, or an angiotensin-converting-enzyme inhibitor as an initial antihypertensive treatment in a diabetic patient as this treatment is known for its protective effect on the kidneys from diabetic nephropathy (Williams et al. 2018).

Controlling hypertension in patients with treatment is a challenge in today’s clinical practice, possibly due to the between-patients’ differences in the aforementioned underlying physiopathologic pathways of the disease’s genesis. A targeted choice of medication might be possible in the future when practical and cost-effective ways of identifying individual mechanisms of hypertension become available to clinical practice. Until then, the pragmatic evidence-based approach is adopted and the main criterion determining the practice is “what reduces mortality in the general population?”. For instance, without scrutinizing patients’ individual specificities, low-dose diuretics have been found to reduce incidence of events and mortality from cerebrovascular diseases, among other cardiovascular diseases (Psaty et al. 2003). This pragmatic approach also allowed the identification of the effect of lifestyle changes on hypertension despite the limited understanding of the underlying physiopathologic mechanisms (Bibbins-Domingo et al. 2010, He et al. 2011, He & MacGregor 2011). While the recommendations are that lifestyle modification should not delay the initiation of the drug therapy, healthy lifestyle has been found to have the
potential to delay or even prevent high blood pressure. High dietary sodium intake, for example, has been associated with a substantial increase in blood pressure values prompting the recommendation of dietary sodium restriction (Elliott et al. 1996). The harmful effects of alcohol consumption on the prevalence of hypertension and the values of blood pressure, on the other hand, have long been established (Cushman et al. 1998). Binge drinking is, thus, highly unadvised, reduction of alcohol intake is recommended in hypertensive patients, and moderation as a mean for prevention (Williams et al. 2018).

In addition, a strong body of evidence has been finding that many diet-related factors, in addition to sodium and alcohol intake, have a significant role in the onset and the progress of hypertension. A diet rich on potassium for example has been associated with lower values of blood pressure in both hypertensive and non-hypertensive individuals. However, the effect of potassium on lowering blood pressure is attenuated with a low sodium intake. That is a bit ironic considering that reduction of sodium intake is advised. For this reason, potassium is recommended at a fixed moderate level while salt is recommended to be decreased (Whelton et al. 1997).

Dietary patterns have also a significant effect on the values of ambulatory blood pressure and the mortality attributed to hypertension. In the early 1960s for example, some regions of Greece and southern Italy have been witnessing some of the highest life expectancies in the world and very low rates of NCDs in general. This status of good health has been linked with their dietary traditions which then became known as the Mediterranean diet. This diet favors fruits, vegetables, nuts, and cereals over red meat, sweets, and dairy products. Fish, poultry, and wine are also consumed in moderation (Willett et al. 1995). The Mediterranean diet has been associated with lower 24 hours blood pressure, lower serum glucose and lipid levels, as well as a significantly reduced risk of cardiovascular events and all-cause mortality (Doménech et al. 2014, Estruch et al. 2018). Its healthy benefit would, in a synergic fashion, be more prominent if accompanied by physical activity and weight loss (Williams et al. 2018).

In fact, excessive increase in weight is directly associated with increased values of blood pressure. A decrease of 5.1 kg in weight, for example, has been found to correspond to a decrease of 4.4 mmHg in systolic blood pressure and 3.6 mmHg in diastolic blood pressure (Neter et al. 2003). To give a better illustration of the significance of this reduction, a 3 mmHg reduction in systolic blood
pressure has been associated with a 5% lower risk of mortality from CAD and 8% lower risk of mortality from cerebrovascular disease (Stamler 1991). Similarly, in a metanalysis of RCTs that studied 5223 participants, significant reduction in systolic and diastolic blood pressure has been associated with different types of physical activity. The reduction in systolic blood pressure/diastolic blood pressure ranged, depending on the type of exercise, from 1.8/2.5 mmHg to 10.9/6.2 mmHg in the general population, and 8.3/5.2 mmHg in hypertensive patients (Cornelissen & Smart 2013).

2.2.3 Type 2 diabetes mellitus

Type 2 diabetes mellitus (T2DM) is a non-communicable disease that has become one of the biggest epidemics in the world with an estimated global prevalence exceeding 380 million adults in 2015 (Zheng et al. 2018). Findings from the Global Burden of Disease Study 2017 attributed an annual death-rate of over a million to T2DM. A rate-increase of 43% since 2007. While these numbers are but to be taken seriously, the effect of diabetes on the worldwide burden of disability is also as big a public health issue due to the slow and silent nature of the disease and its damaging complications. For instance, the Global Burden of Diseases Study estimated that a total of nearly 20 million years lost were to be attributed to T2DM in 2017 (GBD 2017 Causes of Death Collaborators 2018).

T2DM is a disease with an onset that occurs, in most of the cases, long before its diagnosis is established. T2DM complications tend, thus, to have enough time to develop and end-organ damage could reach advanced stages before any clinical manifestation is noted. It is thought that nearly half of the current prevalent cases of T2DM are undiagnosed (Beagley et al. 2014) unnecessarily hindering the burden of the disease and its economic cost which is found to be, by far, underestimated (Zhang et al. 2009). The average healthcare expenditure of a diabetic patient tends, in fact, to be three times more than that of a non-diabetic person (Rubin et al. 1994).

As a trend, the rates of T2DM increased tremendously all over the world over the last four decades. The global number of adult patients living with diabetes raised by 4 folds between 1980 and 2014. This number is projected to increase by 50% in 2035. The burden of this prevalence tends to be heavier in poor countries (Guariguata et al. 2014, NCD Risk Factor Collaboration (NCD-RisC) 2016). Until serious interventions and efforts are taken to manage the disease and address its
development, macrovascular and microvascular complications of T2DM will remain a major public health concern (Kahn et al. 2014).

Historically, the pathogenesis of type 2 diabetes as we know it today started to take shape in the scientific literature since the advent of immunoassay and the identification of beta-cell islet dysfunction (Wu 2006, Yalow & Berson 1960). Prior to that, the variation of response to insulin was already recognized and associated to an unknown condition (Himsworth 1936) which was not identified until the immunoassay breakthrough enabled us to understand the pathophysiology of insensitivity to insulin; nowadays referred to as Insulin Resistance. The scientific community agreed then that this unknown factor responsible of Insulin Resistance can be partially explained by adiposity and serum free fatty acids concentration (Reaven 1988). The accumulation of fat in the intra-abdominal region was then found to be the main location determining the association between adiposity and insulin resistance (Cnop et al. 2002).

In addition to that, it has been found that a feedback loop mechanism is responsible of the late phenomenon of inability to secrete insulin. Feedback loop mechanisms are present in most endocrine systems, and in diabetes, as the beta-cells released insulin acts on insulin-sensitive tissues prompting them to increase their uptake of glucose. These lasts report back to the beta-cells with their glucose needs asking them, through mediators that has not yet been fully identified (Kahn et al. 2014), to raise or reduce their insulin output. When these tissues become less sensitive to insulin, as with obesity for example, their need for glucose rises and beta cells raise their insulin output as a compensation until a point is reached where they are not capable to maintain the glucose tolerance level of the target tissues. The long-term result is thus an increase in serum glucose level. For instance, a study of more than 6000 Finnish men shows that insulin resistance starts its progression long before serum glucose levels start showing abnormal values, and that diabetic dysfunction of beta cells would already be well established before prediabetes could be clinically diagnosed (Stancáková et al. 2009).

Such findings may give the idea that screening for early stages of insulin resistance and beta cells dysfunction could help identifying imminent cases and prevent T2DM. While that can still be relevant, it has been shown that diminished beta cells function can already be present in individuals known to be at risk of T2DM such as patients with polycystic ovary syndrome or gestational
diabetes (Kahn et al. 2014). Moreover, individuals with a family history of diabetes have also been found to have an altered function of beta cells (Cnop et al. 2007). For instance, the heritability of beta cells function has been demonstrated (Elbein et al. 1999) and, according to a study conducted on different ethnic groups in the US, beta cells function was found to have a determining role on glucose intolerance (Jensen et al. 2002). Over a hundred gene loci have then, mainly thanks to genome-wide studies meta-analyses, been linked to type 2 diabetes, serum glucose levels, and insulin concentrations. The majority of these loci tightly influence the function of beta cells (McCarthy 2010, Morris et al. 2012, Scott et al. 2012). However, even combined, these genes only explain a small part of the disease pathogenesis and, if used in predictive modeling, would contribute with but a tiny increase in the prediction of diabetic risk (Meigs et al. 2008).

While the focus of our study is on middle-aged men, it is also worthy to note that the mother’s health behavior, determining the in-utero environment of her offspring which may induce some epigenetic changes, has been found to have a marked effect on the risk of development of T2DM, partly through obesity-associated pathways (Guénard et al. 2013).

Although tending to be a complicated factor, diet, on the other hand, has probably the greatest role in the risk and pathogenesis of T2DM as an environmental factor. Caloric intake for instance and obesity constitute the main predictors of diabetes mellitus through the increase of the predisposition to insulin resistance (Reaven 1988). Specific nutrients and dietary compositions, notably those rich in saturated fat (such as animal fats) and trans-fat (such as hard margarine fat) tend to have a noxious effect on glucose metabolism significantly raising the risk of diabetes mellitus. On the other hand, a diet rich in fibers and with non-hydrogenated polyunsaturated fatty acids, such as those in olive and nut oils, instead of saturated and trans-fat was found to help prevent T2DM (Hu et al. 2001). Magnesium supplementation and diets rich in magnesium such as whole grains have also been associated with lower incidence of T2DM (Salmerón et al. 1997b, 1997a).

A strong body of evidence also explains the role of sedentary lifestyles in obesity and T2DM. Moderate physical activity, even for only 40 minutes per week, has been found to have noticeable benefits in the prevention of T2DM (Lynch et al. 1996). Focus is also rising on the role of sedentary behaviors in determining health outcomes. TV watching for example, which is an activity that has a lower metabolic rate in comparison to other sedentary activities such as reading or car driving,
according to a study of more than 50,000 middle-aged men, increases the risk of being overweight by 4 folds in individuals who watch TV for more than 41 hours a week in comparison to those who watch it for less than an hour a week (Ching et al. 1996). In a recent meta-analysis, the association between TV watching and the incidence of type 2 diabetes was directly examined and a risk increase of 20% was found starting from 14 hours of TV per week with a remarkable linear dose-response relationship (Grøntved & Hu 2011).

Recent studies suggest the existence of other environmental agents, tagged “nontraditional” risk factors, that may be involved in the pathogenesis of diabetes mellitus. For instance, the previously mentioned obesity-associated pathways can be worsened with some environmental chemicals that have been found to cause adverse effects on the metabolism such as disrupting the adipogenesis process. Impairment of this last can promote a state similar to that of lipodystrophic syndromes in which the body is unable to properly store fat significantly raising the risk of T2DM (Auerbach et al. 2016). Further, while recent evidence illustrates the role of inflammation in beta-cell dysfunction and T2DM pathogenesis (Nicol et al. 2013), it is already established that interventions to lifestyle can significantly lower inflammatory markers some of which have been found to correlate tightly with the function of beta-cells (Haffner et al. 2005).

While the focus of this study is on life expectancy or mortality as an outcome, the relevance of our fixation, and that of some clinicians, on diabetes is more due to its role as a risk factor in other diseases mainly the mortality-leading cardiovascular diseases. The presence of a diabetes as a risk factor considerably raises the risk of occurrence of major cardiovascular events and serious damage to end-organs more than most other risk factors especially when co-existing with other comorbidities such as hypertension and risk factors such as lack of physical activity. In such cases the cumulative risk has a negatively synergic effect. For this reason, the presence of diabetes in a patient with hypertension, coronary artery disease, or some other selected diseases, may utterly change the management approach of that disease (American Heart Association 2015, Piepoli et al. 2016). In addition, statistics have shown that even after the occurrence of major cardiovascular events such as myocardial infarction, mortality of diabetic patients was higher than non-diabetic ones (Kiani et al. 2016).
On that account, health behavior, particularly that related to diet, proves to be of an utmost importance in term of T2DM risk prediction. For both prevention and management of current diabetic patients, weight reduction is a priority target component of lifestyle modification, and glycemic control remains a key in slowing the disease and limiting its complications (Kahn et al. 2014). Healthy lifestyle modification was proven effective despite the age category to which an individual belongs. Intensive lifestyle modification was found to be even more effective in reducing diabetes mellitus incidence in individuals who belong to age categories older than young adults (Diabetes Prevention Program Research Group et al. 2006).

2.2.4 Obesity

As all living beings, humans have long had a considerable part in “the war of nature” struggling for existence and running “from famine and death” (Darwin 1859). In past times when the scarcity of food was the norm for most of the common folks and malnutrition the daily bread and butter of medical practice (Adamson 2004), being fat was considered a luxury and a sign of class, wealth, and good health. The status quo did not change until food became more accessible with the generalization of farming and the industrial revolution, and while high corpulence remained socially an indicator of sumptuousness (until probably when strong curves started to run out of cool late 19th century) obesity was not recognized as a disease before it was associated with diabetes, heart disease, and increased mortality with the advances in epidemiology by the half of the 20th century. The overweight trends have also only risen recently, and as a public health problem, obesity in fact dates back to no more than few decades (Eknoyan 2006).

Defined as a condition in which adipose tissue is morbidly accumulating fat to a harmful level, obesity is today a well-established – crucial - modifiable cause of premature death and morbidity. The advent of the obesity epidemic over the past 30 years has been catalyzed by industrialization and driven by rapid changes in populations’ lifestyles and socioeconomic conditions. The epidemic, which some preferred to give the name "globesity", presents today a global challenge to disease prevention and a real public health crisis. Based on the current disease pace and secular trends, projections estimate that in 2030, more than a quarter of the world’s population will be overweight and nearly 15% will be obese (Deitel 2002, Kelly et al. 2008).
Geographically, it was in the USA and Europe where the disease started first to ring the bell of an epidemic in the making. In the US for example, the prevalence of obesity doubled since 1960 hitting the mark of 23% in 1994 and then 32% ten years later before flattening off on that level (Flegal et al. 1998, Ogden et al. 2014). In Europe, the advent of the disease was slower going from 13% in 1992 to 17% in 1998, according to a major multinational European study (von Ruesten et al. 2011). However, as significant regional disparities are noticeable, obesity in some European countries then nearly equaled that of the USA with a prevalence of 31% in the Czech Republic in 2005 (Berghöfer et al. 2008). More toward the north, the share of obese individuals among the Finnish population of adults, according to the Finnish National Institute for Health and Welfare (THL), was of 20.5% in 2018 (Terveyden ja hyvinvoinnin laitos 2018). In the rest of the world, although the epidemic tends to be of a more recent emergence than in the USA and Europe, its depth and spread are showing an alarming acceleration placing some countries beyond the USA’s numbers as in Mexico with a rate of 32.4% of obese adults in 2012 (Barquera et al. 2013).

The challenge of obesity partly resides in its complex and chronic nature as a disease with multiple factors involved in its genesis and upkeep. The consequences of these factors, notably genetic, behavioral, and environmental, gravitate into one pathophysiologic pathway that is modeled by a loss of balance between energy intake and expenditure resulting in weight gain (Garrow 1988). In other words, the origin of this energy imbalance arises beyond the individual’s specifications to partially concern etiologies from the depth of his socio-economic environment (Hruby & Hu 2015).

The model that we are proposing on Figure 3 illustrates some of the interactions between the main categories of obesity etiologies. As childhood obesity is considered to be one of the strongest determinants of adulthood obesity multiplying the disease risk by five, the role of the family in preventing the disease becomes crucial (Simmonds et al. 2016). Moreover, the circumstances of maternity such as weight gain and the occurrence of gestational diabetes during that period have strong effects on offspring obesity (Drake & Reynolds 2010). Adding to that, childhood household, through diet and parental obesity and behavior, has a huge contribution to both childhood and adult obesity (Bammann et al. 2014). For such reasons, family-based approaches are regarded as the best treatment of obesity in children. These approaches act through affecting family and child behavior and have the potential to help prevent obesity in adulthood. It is also noteworthy to mention that 80% of teenagers with obesity tend to carry it to adulthood (Skelton et al. 2012).
Our capacity as humans to store energy into fat is a genetic feature that allowed us to survive throughout the dark and the bright times of history. Some known genetic defects and diseases have a direct relationship in exaggerating this feature such as Cushing’s disease, hypothyroidism, and polycystic ovary syndrome, many of which have a prominent genetic component. Some other defects concern deficiencies of elements that are necessary to regulate weight gain such as in growth-hormone deficiency, a condition in which fat mass is increased in the account of lean body mass. Similarly, deficiency in leptin protein is strongly associated with a morbid increase of body...
weight. However, these conditions tend anyway to be too rare in frequency to explain the obesity epidemic, and while data firmly suggest the implication of leptin protein and its signaling in the pathogenesis of obesity, “leptin resistance” is most commonly a consequence rather than a cause of obesity (Bray 1999, Myers et al. 2010).

Furthermore, even though, genome wide associations revealed more than 60 genetic markers associated with an increased BMI, the common majority of these markers could not explain, even combined, more than 1.5% of the variation in weight (Hindorff et al. 2009). However, the theory that the pathogenesis of obesity is modeled by an interaction between genetic and environmental factors, which is accounted in the genetic susceptibility theory, is a strong one and may explain a bigger proportion of the etiopathogenesis than what has been found so far (Clément & Ferré 2003, Llewellyn & Fildes 2017).

In plain English, hereditary predisposition to overweight does not mean an inevitable fate of living and dying with it. While a family history of obesity reflects that the shared gene pool and the environment may favor it, families can change their environment by promoting a healthy lifestyle and improve their health and that of the coming generations (Centers for Disease Control and Prevention 2018, Harvard THC School of Public Health 2012). What tends to matter beside behavioral and environmental risk factors is their effect in combination to that of the genes under what is called the Gene-Environment Interaction. An example of that would be the work of Qi et al. (2014) which findings’ suggested that some genes only promoted obesity in individuals who with increased intake of fried foods. Similarly, long hours of TV watching were found on another study to enable some genes to promote obesity and physical activity to weaken some other genes’ association with obesity (Qi et al. 2012).

Diet and physical activity tend to share the biggest chunk of the role of behavior in obesity taking the spotlight in most of the studies done on individual’s modifiable risk factors of overweight. Many parameters related to diet affect its relationship with obesity: quantity, quality, frequency, patterns of eating, cooking methods, … etc. Even the fact of just adhering to a healthy diet, no matter its type, has been found to have a protective effect against obesity (Hruby & Hu 2015, Sacks et al. 2009). In similar fashion, adherence to any type of regular activity can have a protective effect
against obesity and if coupled with dietary restriction, these two factors would have a synergic effect (Donnelly et al. 2009).

Other than physical activity and diet, weight gain is promoted by some sleeping patterns, sedentary and television viewing times, as well as any leisure-time sitting (in)activity. Socio-economic conditions have also a strong tie to obesity. In the USA, it was the norm that belonging to the wealthy class is associated with obesity, until about 30-40 years ago, probably because of a change in the sociocultural circumstances of the country with food becoming more available and overweight outmoded. The opposite trend is happening today in the USA and across 11 Organisation for Economic Co-Operation and Development (OECD) countries as overweight became more associated with lower income with a higher inclination to obesity in women. (Devaux & Sassi 2013, Hruby & Hu 2015). This trend comparably applies to education level which seems to have a protective effect against obesity but also against the obesogenic effect of increasing purchasing power. In theory, this offset effect of education can be deduced from observing populations in transitioning economies, but in practice, it might not always be the case (Aitsi-Selmi et al. 2014).

The model is but an oversimplification of what we have found to be relevant etiologies of obesity at a macro level. The fact that some circles are separate does not mean that they are completely independent as factors. The role of policies, societies, and the environment in shaping behavior is well established today and the “nudge theory” is but an example of that (Vlaev et al. 2016). A more systematic way of evaluating the importance of the different factors involved in the disease etiopathogenesis could give a better determination to what etiologies should be put in the model, and a weighing process could set the size of the circles and provide the model with a proper academic relevance. Appropriate addressing of the disease causes would necessarily involve multiple disciplines. A systems-oriented approach might be the best way to address the complexity of obesity. Such approach allows the creation of a unified theoretical framework enabling multiple disciplines elaborating at different levels to effectively work on the same problem (Huang et al. 2009). Although that may sound futuristic for public health, bottom-up approaches are not that new and have been used in different ways and for different purposes (McLeroy et al. 1988). They might be relatively hard to put to practice, but experiences from different regions showed that they are
not impossible to implement. The North Karelia project might be a good example to mention here as a successful bottom-up approach to address obesity (Lindström et al. 2016).

Before getting to the consequences of obesity, it is worthy to mention that the harmful effects of the disease are not only function of the amount of weight gained but also on the nature of the built-up mass and the patterns of fat distribution. Excess abdominal fat for example is known to favor T2DM and to be of higher cardiovascular risk than other body “builds” and excess body fat in general (Cnop et al. 2002). Hence, the concern with obesity goes beyond excessive overweight and, in addition to the metabolic circumstances and consequences of the disease, encompasses body shapes and fat distribution.

The field of clinical anthropometry tries to concentrate on the study of various body proportions and dimensions and their effects on health. While nowadays methods of assessment of body builds and fat proportions vary from skinfold calipers and body circumference measurements all the way to 3D Body Scanners, Bioimpedance Spectroscopy, and Multi-Compartment Models, clinical practice and non-obesity focused research requires simple tools that are able to give information on an individual’s weight status in an efficient, yet relevant and significant manner. For this reason, the most widely used measures to classify body weights in adults are waist circumference and Body Mass Index (BMI). The latter is an indicator of the weight over height ratio and ranges from underweight (BMI < 18.5 kg/m²) to severe obesity (BMI ≥ 40 kg/m²). In adults, a BMI below 18.5 indicates underweight; normal weight is between 18.5 and 25; 25 to 30 is overweight and anything beyond is obesity. In children, BMI is used in comparison to reference tables such as the WHO Child Growth Standards taking into account individuals age, among other parameters (Hruby & Hu 2015, Keys et al. 1972). Whether BMI is the best measure of adipose tissue abnormal distribution and the most accurate indicator of the risks that obesity raises remains a controversial question. However, as Adab et al. (2018) articulated it: “It works for most people most of the time”. Hence our decision to use it to measure obesity in our study.

As of its consequences, in overall, it is estimated that excess weight deteriorates life expectancy by 4 to 7 years (Peeters et al. 2003). For example in the USA in 2000, overweight and obesity were found to be responsible of 15% of the mortality attributed to sedentary lifestyle and unhealthy diet (Mokdad et al. 2004). Obesity is a common risk factor for most non-communicable diseases. It
contributes in the pathogeny of diabetes, cardiovascular (Flint et al. 2010) and cerebrovascular diseases, and raises the risk of certain cancers (Polednak 2008), osteoarthritis, mental disorders and many other diseases. For instance, obese people have seven times more risk to have diabetes in comparison to people with normal weight (Abdullah et al. 2010). Hence the rule in the USA to screen anyone over 45 years of age with excess weight (American Diabetes Association 2012).

Many surgeries also tend to get riskier with excess weight. The risk of major complications of surgery is double for example in obese trauma patients in comparison to normal weight trauma patients (Glance et al. 2014). With excess weight, hospital stay tends also to be longer and the complications associated to it, such as surgical-site infection, nosocomial infections, and pressure ulcers, tend to be higher in both frequency and severity (Huttunen & Syrjänen 2013).

Economically obesity causes unmediated burden through employees’ performance, sick leaves, and longevity. As of health expenditure, it was estimated that, in Europe, a quarter of medication costs is generated by excess weight. Obese individuals annually incur up to 1800 EUR excess in direct costs of care in comparison to their normal weight peers. Health systems are billed nearly 5% of their total healthcare costs because of excess weight (von Lengerke & Krauth 2011). In Finland for example, it was estimated in 2006 that the government spent over 260 million EURO in direct obesity-related costs. In addition, weight-loss programs were supported by 800,000 EUR from the Finnish Social Insurance Institution (Kela) (Baker 2011).

Obesity today became a heavy disease. Heavy by its health burden and economic cost, but also heavy by its complexity and difficulty to tackle as a public health problem. Its consequences are to be taken beyond the concern of the individuals but as a global scale problem. For instance, obese people tend to be almost 9 times likely to fail to acquire the immune effect conferred through vaccination than non-obese individuals (Young et al. 2013). The advent of obesity in a population should raise concern regarding the robustness of its vaccination herd effect.

Obesity is caused by a myriad of etiologies interacting with each other. After thoroughly understanding them, addressing the epidemic requires to act on the powerful forces, notably behavioral and biological, that are modeling the disease. The interaction between individuals and
their socio-economic context imposes for the solution of obesity to operate on the individuals’ behavior in respect to their environment and circumstances.

2.2.5 Cancer

After cardiovascular diseases, cancer is the next major responsible of global mortality. The cancer-attributed proportion of deaths exceeded the sixth of all deaths in 2018 with a toll nearing 10 million (World Health Organization 2018c). The lifetime risk of developing cancer – all kinds included – is about 47.5% in the UK and 40% in the USA (American Cancer Society 2018, Smittenaar et al. 2016).

Cancer is not a single disease, but a term referring to hundreds of malignant conditions similar in their invasive and metastatic nature. Each condition is unique in its biological characteristics, risk factors, etiologies, pathogenesis, diagnosis approach, and treatment. The technological advances in imaging and analysis tools, as they uncover further subtypes, are but adding to the complexity of this conglomeration of diseases. Consequently, beside that the claims announced every now and then - and going viral - promising that a cure to all cancers is coming up in a year are very unlikely to meet reality, an explanation of cancer’s epidemiology through one or few of its forms would not be realistic neither. Therefore, the solution to this most challenging global health issue might not be an elixir.

A considerable part of prevented cancer cases and prolonged survival are attributed to public health efforts and technological advances. Death rates from cancer, for instance, tend to be higher in low and mid-income countries the majority (80%) of which do not possess proper data systems to optimize the public health management of the disease (World Health Organization 2018c). And although survival rates have been remarkably improving in high-income countries (Fang et al. 2013), historical rates have been much better than today (Figure 4).

Lung cancer is the form of cancer with the highest global incidence and the highest annual number of deaths with 1.76 million deaths in 2018 (World Health Organization 2018c). Its particularity of being known to have a strong and clear association with smoking draws the attention toward the prevention of exposure and the role of behavior.
The graphs suggest that mortality rates attributed to lung cancer were much lower in the 1950s than since 1970. The graphs also show a gender-gap in lung cancer mortality shrinking over time. Considering the historic lag of smoking adoption between females and males, the delay of increase of female mortality rates could be explained by the difference in the development of their smoking patterns in comparison to men (Figure 4). Also, knowing that males and females in the same country usually benefit of similar diagnostic capabilities, improvements in diagnostic technologies might not be the explanation of the rise of mortality over the decades and the role of the behavioral and environmental factors in the disease come to the foreground once again.

This trend of gender-gap, lower historical mortality rates, and recent regression seems to be affecting different countries in a similar fashion suggesting an effect of the globalization of behavioral trends as well as the advances in healthcare and prevention. For instance, while the
highly fatal nature of lung cancer imposes that its mortality rate almost equals its incidence rate, trend studies have shown that a significant increase in lung cancer incidence appears approximately 20 to 30 years after a population is hit by the tobacco smoking epidemic (Thun et al. 2012).

Migration studies, on the other hand, have found that second and third generation migrants tend to have their incidence rates shifting in the direction of the incidence rates of the hosting population suggesting that behavioral and environmental risk factors are of great etiopathogenic importance (Schottenfeld & Jr 2006). Epidemiologic studies have demonstrated that 9 out of 10 of all lung adenocarcinomas can be actively avoided. Occupational hazards, air quality, ionizing radiation, and diet explain a part of the etiopathogenesis but their role tend to be neglected in comparison to tobacco smoking (Office of the Surgeon General (US) & Office on Smoking and Health (US) 2004). Also, regardless of cancers’ forms and etiologies, it is an established fact today that more than 30% of deaths attributed to cancer in general are preventable and could be avoided through behavioral change and lifestyle modification (World Health Organization 2018c).

One of the reasons cancer remains a challenge to modern medicine is the complexity of the molecular nature of the disease. Pathologically, all cancers tend to arise from alterations to the genetic code of cells causing a disruption of their balance and some of their main functions. Consequently, the cells proliferate in an unreasonable manner, forming into tumors, invading proximal tissues, and metastasizing distantly. The environmental and behavioral risk factors have in common that they all tend to raise the frequency of these mutations and cause the normal genes to lose control of cells (Garraway & Lander 2013). The main cellular functions which the disruption might lead to cancer are linked to a number of genes called driver genes which the mutation might contribute in the genesis of cancer (Bailey et al. 2018). These genes are categorized into: Proto-oncogenes code for proteins that accelerate cell proliferation. Tumor suppressor genes are meant to limit proliferation and, in most cancers, are either absent or deactivated. DNA repair genes are meant to proofread genetic codes and eliminate most of the seldom-occurring errors, a function that often gets deteriorated in cancer cells. Pro-apoptotic genes, on the other hand, are meant to promote programmed death of cells and death after damage (Vogelstein et al. 2013). Luckily, and unluckily, many of these mutations need to co-occur simultaneously for cancer to arise. Luckily:
because that makes cancer formation a relatively rare phenomenon, and unluckily: because targeted
treatment approaches are likely to fail curing an already established cancer (Benz 2017).

While chemotherapy remains the widely used treatment for the majority of cancers, advances in
targeted therapies allowed the rise of a generation of drugs such as the ‘smart bombs’ that target
specific molecular dysfunctions based on intelligent drug delivery systems (Behrooz et al. 2017).
These drugs along with immunotherapy, and precision cancer medicine constitute the aspired
strategy to control the cancer epidemic in the future. However, making these treatments sustainably
available at a global scale would probably be a greater challenge than advancing the very
techniques. A widely affordable precision medicine, for example, is a dream unlikely to see
daylight in the near future; another reason to shift the approach of cancer management in public
health to focus more on prevention rather than on survival (Benz 2017).

Like tobacco smoking, the role of obesity, insulin resistance, and other risk factors in cancer are
well established today (Gallagher & LeRoith 2010, La Vecchia et al. 2011). Like the previously
discussed comorbidities, the roots of the risk of many forms of cancer can, partly, be traced back
to health behaviors. This root linkage through health behaviors could elucidate some hard-to-
explain connections between the comorbidities. Cardiovascular risk, for example, is significantly
increased by cancer in cancerous patients with traditional cardiovascular risk factors. Their
cardiovascular risk remains increased even after successful treatment of their cancer with
chemotherapy or radiotherapy (Piepoli et al. 2016).

A growing body of research is supporting the impact of cancer prevention strategies and
community prevention interventions on the reduction of cancer incidence and mortality (National
Cancer Institute 2017). These interventions must be tailored to fit different ages and genders as
well as the individual specific environments and risk factors. Adults at midlife tend to cumulate
more risk factors than younger individuals, but they are more likely than elderly individuals to
benefit from behavioral changes (Ory et al. 2014). Chemoprevention opens new possible
perspectives to cancer prevention (Siddappa et al. 2017, Yates et al. 2015), but in the current
circumstances, from a public health perspective, promoting health-behaviors can potentially yield
more substantial results.
2.2.6 Comorbidity evaluation

Different methods and scoring indexes have been developed to evaluate multimorbidity and comorbidities. Some studies are based on counting the number of comorbidities and used the International Classification of Diseases (ICD) to measure comorbidity such as Cho et al. who used ICD-9 to identify comorbid conditions and then to create a comorbidity score for which estimation of life expectancy was adjusted (Cho et al. 2013). Such ICD based methods might be lengthy and difficult to realize and, as we mentioned, would limit multimorbidity estimation on the number of diseases in an individual regardless of the severity of their concurrence. Another type of methods considers both the concurrent presence of diseases and their consequent cognitive and physical functional disability. This method might be a bit too complicated for clinical studies. However, cumulative indices methodologies, another type of methods, according to Marengoni et al. who explained these three different perspectives of addressing multimorbidity, could be, in term of complexity, placed between the two previously mentioned ones (Marengoni et al. 2011).

An important one would be the Cumulative Illness Rating Scale (CIRS) proposed by Linn et al. (1968). This score includes in its evaluation of comorbidities, all possible primary care problems and an estimate of the gravity of different combinations of chronic diseases through an organ-wise weighing system (Linn et al. 1968). The tool has been documented to have a good interrater reliability. This last is a validity test that scores the consistency of rating given by different judges. CIRS has also been found to have a high predictive validity for mortality (Hudon et al. 2005).

A simpler cumulative score that is also widely used to evaluate comorbidities would be the Charlson Comorbidity Index (CCI). This method relies on ICD diagnosis codes to categorize morbidities and gives them a weight from 1 to 6. The final score takes into account the effect of the coexistence of the weighed comorbidities and would represent the risk of mortality and resource use (Charlson et al. 1987). Since its conceptualization in 1987, there was many attempts to improve the use of the index. Deyo et al. (1992) reduced the score’s count of categories from 19 to 17 and helped its adaptation to the Clinically Modified 9th Revision of the International Classification of Diseases (ICD-9-CM). Comorbidities’ weights have also been modified (Schneeweiss et al. 2003) and with the release of ICD-10, Quan et al. (2005) adapted the score to its new coding system.
The CCI, widely used and validated, has been proved to be an excellent determinant of comorbidity status for outcome prediction. However, CCI is limited in the sense that it does not consider the severity of most of the present comorbidities. For instance, a patient with chronic obstructive pulmonary disease (COPD) would receive an additional one point to his CCI score no matter the severity of their COPD. Usually determined in accordance to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines, the severity of COPD is clinically assessed based on symptomatology and spirometry. The GOLD grade classifies COPD patients into four stages of severity. The milder of the stages typically differs significantly in mortality from the most severe (Leivseth et al. 2013).

On the other hand, the opposite can also happen. A patient with a diagnosis of HIV/AIDS receives a score of 6 in CCI indicative of a severe comorbidity burden. However, a good management of the disease is possible nowadays, although still function of the location and probably the socioeconomical conditions. It is argued that the 1987 weighing system of CCI is losing relevance with time and shall be reviewed (Holmes et al. 2014).

Another comorbidity determinant that is used as an adjunct to outcome prediction is comorbidity-polypharmacy score (CPS): a simple clinical tool that has been proposed in 2011 (Evans et al. 2011) as an index to measure the “cumulative severity” of comorbid conditions and improve outcome prediction, especially in trauma patients. Polypharmacy, refers to the simultaneous use of different drugs, is used in CPS as a surrogate of the intensity of treatment required to control the present comorbidities, and thus, can play the role of an indicator of the severity of these lasts. In addition, polypharmacy is associated with major risks related to medications side effects, interactions, usage errors, and nonoptimal adherence to medication.

Easy to calculate, CPS simply counts the number of comorbidities and medications used assigning one point for each comorbidity and one point for each of the patient’s medications with no maximum score. Patients are then classified by severity as mild (CPS 0-7), moderate (8-14), severe (15-21), and morbid (≥22 points) (Evans et al. 2012, Stawicki et al. 2015).

Results from research that examined how CPS predicts morbimortality were inconsistent. Nossaman et al. studied 446 middle age and older trauma patients and examined the predictive
effect of both CCI and CPS on mortality as well as hospital stay. They controversially found that, in older patients, high CPS is associated with lower mortality. The authors commented that this paradoxical effect, which was found with CPS and not with CCI, could possibly be explained by a positive correlation between high medication and appropriate care and diagnosis (Nossaman et al. 2018).

Comorbidities and their severity are considered significant confounders and thus constitute a very important source of bias in observational studies, especially those related to survival. Ignoring their confounding effect might rise falsely significant differences (Fouche et al. 2017). While CPS proposes better insights on disease severity, its use is more appropriate in trauma settings, and it lacks validity in term of chronic disease survival prediction when compared to CCI.

Although some studies suggest that Age-Adjusted Charlson Comorbidity Index (AACCII), a modified version of the Deyo et al. (Deyo et al. 1992) Charlson Comorbidity Index score which includes age as a correction variable (Koppie et al. 2008, Robbins et al. 2013) might provide better insights in clinical studies (Yang et al. 2015).

2.3 Risk factors in health

2.3.1 On risk and causation

It was common in antiquity and early middle ages for Europeans kings to consult oracles and fortune-tellers regarding matters related to upcoming threats. Alexander the Great consulted Pythia when he wanted to know about the future in a time when science and philosophy were thriving among the Greeks. The common thought back then was that analytical thinking had no relation with the future (Bernstein 1998).

The concept of risk fundamentally involves the future. While the term risk originates from the Arabic word رزق (rizq) and the term hazard from the Arabic word الزهر (alzahr) - the former referring to an upcoming good fortune and the latter to fortunate luck -, the Arabs could only dwell with the concept of risk in their theoretical writings. In times where algebra was thriving in Baghdad, through risk calculation, mathematics was considered the keyhole through which we can look into the future (al-Baghdadi 1020). However, beyond basic predictions of armies’ logistical needs, the Arabs were not able to employ their reflections about risk into practice. Their theories,
algebra, and numeric systems might have found their adaptation further in time. One example would be the works of G. Cardano (Cardano & Wilks 1663)– the physician who, in the beginning of the Italian renaissance, tried to employ early mathematical laws of probability to predict the outcomes of his gambling games. The concept of risk was then adapted to many different circumstances, notably in the context of sea exploration when there was a need to estimate the risk of missing in sea and the chance of coming back with gains in order to provide sailors with some sort of insurance and convince them to go in risky sea adventures (Bernstein 1998).

In the context of modern epidemiology, we talk use the term risk when referring to the chance of occurrence of a disease or a health event and risk factors when referring to circumstances that increase the likelihood of a disease or an event to occur. Risk factors are not to be muddled with disease pathogenesis since that this last might be the domain of laboratory-based research and could lag for decades behind epidemiological findings without altering them when established. For instance, while the pathogenesis of sudden infant death syndrome (SIDS) remains misunderstood, some of the disease’s risk factors, notably sleeping on stomach, have been identified (National Institute of Child Health and Human Development 2017) and it is sufficient for public health to have evidence that removing a risk factor improves health to prompt action on it. The decline of SIDS and the success of Safe to Sleep Campaign portrays a good example of this concept (Dwyer & Ponsonby 1996).

In that context, epidemiology may appear to be dealing with causation or causality. However, although the presence of some signs might be able to suggest that an association constitutes a cause-effect relationship (Hill 1965), since most of its studies are observational studies, epidemiology merely indicates association. Risk factors do not represent causation or causality, even though they may contribute to it; they are defined as variables associated with an increased risk of disease but they can be seen as surrogates of components in the sufficient-component cause model as defined by Rothman (1976). In the 1960s for example, when Finland had some of the highest rates of mortality from coronary artery disease in the world (Keys 1970), from an international perspective, speaking Finnish was to be considered as a risk factor of cardiovascular disease. The language was not a component of the cause of the disease but it reflected environmental, genetic, or behavioral components that may favor the occurrence of the disease (Hopkins & Williams 1981).
If we start from mortality and trace back to its etiologies and the causes of its etiologies, we first find that chronic diseases, as discussed previously, are directly linked to a large proportion of mortality globally and an even larger proportion in developed countries. The most important of these chronic diseases are inter-connected in their etiopathogenesis and are modeled by a set of risk factors many of which are shared between these chronic diseases (GBD 2017 Causes of Death Collaborators 2018).

Different classifications tried to sum up risk factors. The use of different groupings can be factor of practicality and context such as the classification into modifiable and non-modifiable risk factors in the assessment of risk in coronary artery disease. In our context, we procure the following classification from the Global Burden of Disease study (GBD 2017 Causes of Death Collaborators 2018) and WHO’s Global Health Risks report (World Health Organization 2009). We group health risk factors into four categories:

- Behavior and lifestyle-related risk factors
- Individual (involves genetic factors and physiologic factors)
- Environmental and occupational risks
- Socio-demographic

WHO stated that 60% of the disease burden in Europe is to be attributed to seven leading risk factors. These factors are either related to lifestyle: tobacco, alcohol, low fruit and vegetable intake, or to metabolism: overweight, high blood cholesterol, high blood pressure (World Health Organization 2006). The latter are often modeled by the former risk factors.

2.3.2 Health lifestyles

To provide an academic explanation of what constitutes a health lifestyle, we derive a definition from the words of Saint Onge & Krueger (2016). Health lifestyles refer to the orientations of behavior reflecting intentions directed toward gaining health in accordance with current knowledge of what is healthy and what is unhealthy.

As we have discussed earlier, the origins of many comorbidities’ etiopathogeneses are dominated by lifestyle-related risk factors (McGinnis & Foege 1993). Eliminating obesity, unhealthy diets,
and physical inactivity could prevent up to 80% of heart disease, stroke, and diabetes mellitus (World Health Organization et al. 2011). In the famous study by Mokdad et al. (2004) “Actual causes of death in the United States”, 50% of the United States’ mortality rate of 2000 could be attributed to smoking, physical inactivity, alcohol consumption, unhealthy diet, some microbial and toxic agents, as well as a few other preventable factors such as car crashes, firearms, drug use, and risky sexual behaviors.

The Global Burden of Disease study, which has involved over 3600 researchers, has been allowing the assessment, among other things, of the causes of premature death in 195 countries since 1990 until today. Diet was considered the main contributor to premature mortality for all ages worldwide from 1990 to 2017, with diets high in sodium and low in whole grains and fruits topping the dietary factors. Tobacco smoking ranked second as a cause of premature mortality among all the behavioral risk factors. When all ages are confined, this last statement would apply to the period from 1994 to 2017. Prior to 1994, malnutrition was the second cause of mortality among behavioral risk factors affecting mainly infants and children. As if 2017, in countries with a low Socio-demographic Index (SDI), malnutrition remained the main cause of premature death. Whereas, dietary factors and tobacco topped the behavioral causes of mortality in countries with middle to high SDI (GBD 2017 Causes of Death Collaborators 2018).

Like other high SDI countries, in Finland, dietary risks and tobacco smoking were the main contributors to premature death followed by alcohol use and low physical activity all along the period from 1990 to 2017 (Figure 5).
Diet

As discussed earlier, dietary habits and nutrition have an important role in the pathogenesis of many non-communicable diseases determining their risks and influencing their attributed mortality. According to data from the Global Burden of Disease Study, dietary risks are found to be responsible of the largest part of mortality attributed to behavioral factors in many developed countries, notably Finland (Figure 5) and the USA (GBD 2017 Causes of Death Collaborators 2018).

From a global perspective, among the dietary risks, high sodium intake has been the number one contributor to mortality. The number of deaths attributed to increased dietary sodium mostly (85%) concerns the association with this last with cardiovascular diseases, notably through the hypertension pathway. The mortality from cancers attributed to high sodium intake only constitute 10% of the total mortality attributed to high sodium intake followed by mortality from diabetes and kidney disease (4%) (Figure 6).
Figure 6 Global mortality from dietary risks
Visualization tool provided by the Institute for Health Metrics and Evaluation (IHME) (2015)

As previously discussed, the association between blood pressure and sodium intake is, in fact, well established today (Elliott et al. 1996). However, although dietary sodium restriction has been found to reduce absolute blood pressure numbers and potentially help reduce the necessary dose of antihypertensive treatment to achieve blood pressure control, the beneficial effects of this lifestyle change were noticed to tend to diminish over time creating doubt on the reliability of salt restriction as an intervention to reduce blood pressure. This could be partly due to the difficulty to persist on that behavior and partly because the relation between salt reduction and mortality might follow a J-curve with a paradoxical effect of increased mortality below a certain level of sodium intake (Mente et al. 2016, Williams et al. 2018).

In Finland, mortality from dietary risks is rather topped by a reduced dietary intake of whole grains, nuts and seeds, also with cardiovascular disease as the main pathway of mortality (Figure 7). The benefits of diets rich in fibers and whole grain have been the subject of many studies for the past...
20 years and were found to be associated, not only with lower risk of obesity, but also with lower risk of heart disease, T2DM, and colon cancer (Ye et al. 2012).

Figure 7 Mortality from dietary risks in Finland
Visualization tool provided by the Institute for Health Metrics and Evaluation (IHME) (2015)

The complexity of human food makes it, from one side, difficult for research to determine the effects of single diet components on health (Willett 1998). From another side, assessing how healthy are an individual’s dietary habits is not an easy task, let alone a population’s (Vafeiadou et al. 2012). Diets are subject to major variations in frequencies, quantities, and compositions depending on subjects, ethnicities, socio-economic and education levels, and geography. Often, the same type of food, for example, vegetable or fish, is found to carry different composition of nutrients and minerals depending on the geographical region. Also, the same nomenclature can refer to a different food in different regions. Different combinations of foods also tend to interact differently with each other, creating patterns of diets. Diet quality scores represent a tool that permits to overcome these problems by directly measuring how healthy is an individual’s diet in reference to a predetermined diet that showed evidence of correlating with positive health outcomes all while going in accordance, for some, with the dietary habits of the region (Waijers et al. 2007).
Waijers et al. (2007) found four major diet scores that have been properly validated and from them many other scores have been derived. These four diet scores are: the Diet Quality Index (DQI) (Patterson et al. 1994), the Healthy Eating Index (Kennedy et al. 1995), the Mediterranean Diet Score (Trichopoulou et al. 1995), and the Healthy Diet Indicator (Huijbregts et al. 1997).

These scores and their derivatives differ in the kinds of foods that compose their healthy reference, and also, partly, in how they include subtypes of fat in their calculation. Monounsaturated and polyunsaturated fatty acids are known to be associated with a reduction of cardiovascular risk, while trans fatty acids are associated with an increase of that risk (Oh et al. 2005, Solfrizzi et al. 2005). It is, thus, important for a score to benefit of these relatively recent findings - and others - and include relevant variables in their calculation. Also, worth noticing, that what would be referred to as brown bread, for example, in one region would indicate a bread with a different composition in other regions. Plus, studies have shown that region-specific dietary patterns differ significantly even between societies with relatively similar culture and values (Freisling et al. 2010).

The Baltic Sea Diet Score (BSDS), also sometimes referred to as the Nordic diet, can represent a healthy diet index that emphasizes on foods that are easily available, affordable, and culturally acceptable in the region of the Baltic Sea and Nordic countries, creating locally relevant standards of what could constitute a healthy diet and what food amounts are to be recommended. The score is developed based on the dietary patterns of a Finnish population and is based on a pyramid model of what is considered to be healthy among the region’s foods based on evidence and recommendations of nutrients that are required to be promoted for the purpose of improving public health in Finland. Foods that are recommended to be consumed the most, such as fruits, vegetables, berries, and roots, are situated at the base of the pyramid while foods that are considered unhealthy such as processed meat, butter, and sweets, top the pyramid. These foods are categorized into ten groups and, using the pyramid as a guideline and based on population consumption quartiles, sex-specific score points are assigned to each consumed component. Consumption of unhealthy foods reduces the score. Consequently, the higher the score, the higher the adherence to the healthy Nordic diet, the better the health outcomes (Kanerva et al. 2014b).
The Baltic Sea Diet Score has been assessed and used in different studies (Isanejad et al. 2018, Kanerva et al. 2013). However, since it is a relatively new score, bigger studies might be needed to support its validation (Isanejad et al. 2018, Kanerva et al. 2014a, 2014c, Maukonen et al. 2016).

**Tobacco Smoking**

Only saying that tobacco consumption constitutes a major cause of global mortality would probably not do justice to the magnitude of the issue. With more than a billion current smokers worldwide, around 6 million deaths are accountable to tobacco smoking every year (Drope et al. 2018).

Global smoking trends have generally been declining. The prevalence of daily smoking in 1980 was estimated at around 41% in men and 10% in women. These values diminished to 31% in men and 6% in women. However, due to the growing number of the population of humans, the absolute number of smokers is rather increasing (Ng et al. 2014).

These prevalence rates vary extensively between the countries. In men, the rates ranged, in 2012, from more than 50% in Armenia, Russia, and Indonesia – among other countries – to below 10% in Nigeria, Ghana, and Sudan – among other countries. In some countries such as Greece and Bulgaria about a third of the population of women were smokers in 2012. On the other hand, in some other regions, it was very rare for women to smoke with rates below 1% in Algeria, Libya, Morocco, and Cameroon – among other countries (Ng et al. 2014).

While the global average of daily cigarettes smoked is of about 18 cigarettes per smoker a day, some regions manifest heavier consumption than others with an individual daily smoking rate of less than 5 in some countries such as Bangladesh, Bolivia, Burkina Faso and Benin and over 30 in others with 39 cigarettes a day per smoker in Antigua and Barbuda, 49 cigarettes a day per smoker in Brunei, and an extreme 109 daily cigarettes per smoker in Suriname (data from 2012) (Ng et al. 2014). However, it has been proved that smoking even just one cigarette a day, which is thought to be relatively harmless, is just as dangerous as heavy smoking according to a solid meta-analysis. "Smokers should aim to quit instead of cutting down" the authors concluded (Hackshaw et al. 2018).
Similarly, waterpipe – also called argileh or sheesha – tobacco smoking was found to be as deleterious to health as cigarette smoking (Akl et al. 2010). Electronic cigarettes, also called e-cigarettes or vape, on the other hand, marketed first in China in 2004 and became widely used since the past 6 years, are thought to be safer than cigarette smoking (Goniewicz et al. 2014). However, the practice is too recent for epidemiological studies to determine its long-term effects. In addition, the chemicals dwelled through the device are not fully assessed, and they tend to vary from manufacturer to another. E-cigarette could constitute a smoking cessation pathway for some heavy cigarette smokers, but the device could cause nicotine addiction in non-smokers. The effective impact on public health of the global spread of the e-cigarette phenomenon is obscure (Harrell et al. 2014, Rom et al. 2015, Schraufnagel et al. 2014).

People usually start smoking in teenage years prompted by psycho-social circumstances. Most of these teenagers, by the age of 20, also regret that they became addicted. Actually, most of the smokers in general regret ever starting to smoke. The psycho-social prompters of the practice usually only concern its beginning, then the pharmacological effects make it a vicious habit that is hard to quit. Nicotine is the substance responsible of addiction in smoking, playing the role of a psychomotor stimulant that induces dopamine release, which stimulates the brain and improves its performances. With time, tolerance to this psychoactive effect develops and cigarette use will only be calming nicotine’s withdrawal symptoms without providing the performance improvements. Nicotine withdrawal symptoms include dizziness, depression, anxiety, irritability, trouble sleeping, feeling of boredom, attention and concentration issues, headaches etc. These withdrawal symptoms begin few hours after smoking and require, for some, up to a month to fully resolve. After this period, complete smoking cessation becomes possible, but the behavioral ritual of the habit might lead to a relapse. Serious abstinence might take years of struggle to be reached (Jarvis 2004, Nayak et al. 2017).

Smoking is known mainly for being the principal cause of lung cancer, responsible for 60% to 90% of lung cancer incidence. Lung cancer is responsible of the highest yearly number of deaths from cancers worldwide, and because it is mainly caused by cigarette smoking, it is considered the most preventable cause of cancer death (The International Agency for Research on Cancer 2004).
The main cause of harm in cigarettes originates from the physical and chemical characteristics of
the cigarette smoke which holds over 3500 different types of particles such as hydrocarbons,
dioxins, furans, heavy metals, and nitrosamines. Many of these particles have carcinogenic
characteristics. Like e-cigarettes, the composition of cigarettes vary from manufacturer to another
and, thus, their carcinogenic effect may vary as well (Stabile et al. 2017).

Because this carcinogenic effect is mostly due to the mainstream aerosol that tobacco cigarette
smoking generates, secondhand smoke causes harm to nonsmokers, notably children (40%), and
an increased risk of respiratory disease, heart disease, stroke, and lung cancer. Secondhand
exposure to tobacco smoke can be measured through cotinine: one of the resultants of nicotine
metabolism and a surrogate that can be measured in saliva, urine or blood. In the United States,
through the measurement of cotinine, exposure to secondhand smoke in nonsmokers declined from
88% of the population in 1988-1991 having significant cotinine levels in their bodies to 25% in
2011-2012. Globally, the number of deaths attributed to secondhand smoking in nonsmokers was
over 600 000 deaths in 2004 (CDC’s Office on Smoking and Health 2017, Öberg et al. 2011).

Cotinine measurement has recently helped prove, through a population-based study, a direct dose-
response pro-inflammatory toxic effect of smoke exposure on the human body (Choi et al. 2019).
For instance, the harmful effects attributed to smoking are mostly caused by a reaction between the
smoke particles and the tissues causing damage to biological structures and processes through
oxidative stress and inflammation, among other mechanisms, impairing repair, initiating
atherosclerosis, and prompting dysplasia (Goldschmidt-Clermont et al. 2012, Sasco et al. 2004,
Tuder & Petrache 2012). The consequences constitute a panoply of diseases going from orobuccal
cancers (Allam et al. 2011, Maruccia et al. 2012), respiratory tract cancers (Sasco et al. 2004,
Semenzati et al. 2012), lung cancer, chronic obstructive pulmonary disease (Tuder & Petrache
2012), gastroenteric tract cancers (González et al. 2003, Sasco et al. 2004), cardiovascular diseases
(Goldschmidt-Clermont et al. 2012, Pipe et al. 2010), all the way to diseases of the kidney, urinary,

In Finland, mortality rates attributed to first-hand smoking are not so different from the global rates
with nearly 100 deaths per 100,000. Most of these smoking-caused deaths are attributed to cancers
(44%) and cardiovascular diseases (33%) followed by strokes and chronic obstructive lung disease (Figure 8) (GBD 2017 Causes of Death Collaborators 2018).

The previously mentioned diseases create a heavy economic burden which is certainly more difficult to bear by low income countries. Nevertheless, these countries are thought to be enduring 40% of the healthcare costs attributed to tobacco smoking. Globally, it is estimated that in 2012, 5.7% of the total health expenditure went to diseases caused by smoking and that, if we include the lost productivity due to the burden of smoking, the total economic cost that the world carries because of this behavior nears 2% of the annual global gross domestic product (Goodchild et al. 2018).

Alcohol drinking
The term alcohol became widely associated with the effects of the short term and long-term use of beverages containing the chemical compound ‘alcohol’. The harmful consequences of these effects on the individual and on the society have been proven and the burden of alcohol has been assessed both at a global scale and on different countries.
Alcohol refers mainly to the chemical terminology of organic compounds that has a hydroxyl group bound to a saturated carbon atom (McNaught & Wilkinson 1997). Alcohol beverages are the drinks that contain ethanol, a type of alcohol that is informally referred to when talking about alcohol in general. More than half of the world’s adults (2.4 billion drinkers in 2016) drink alcohol beverages, making this substance the most commonly used recreational drug worldwide (Griswold et al. 2018) with a prevalence of drinking of more than 87% among the United States’ adults according to the National Institute on Alcohol Abuse and Alcoholism of the USA (National Institute on Alcohol Abuse and Alcoholism (NIAAA) 2018).

‘Alcoholism’ is one of the consequences of the use of alcohol and it is defined as a medical condition caused by the harmful use of alcohol. It involves alcohol dependence and alcohol abuse and it is also referred to as Alcohol use disorder (AUD) (American Psychiatric Association 2013).

Humans consumed alcohol since the dawn of ages. A study that analyzed antique jars, dating from up to 10,000 BC and originating from a Neolithic village in China, discovered traces of distilled conserved alcohol (Patrick 1970). Another study proved that an alcoholic drink was produced through fermentation in 7000-6650 BC (McGovern et al. 2004).

Alcohol was known for having both benefits and harms, as described in the Hindu Ayurvedic scriptures from back to around 5,000 BC. These texts warned of the disastrous consequences of the excessive consumption of Alcohol (Dasgupta 2012). The Quran, the holy book of Muslims, stated as well around the year 620 AD, concerning alcohol and gambling: “In them is great sin and [yet, some] benefit for people. But their sin is greater than their benefit.” (Saheeh International 1997).

However, the health benefits of alcohol are very negligible in comparison to its harmful consequences and for that, health professionals should never, in any case, suggest to patients to drink alcohol (Jackson et al. 2005).

The World Health Organization describes three mechanisms of health damage due to alcohol consumption. The first is related to the long-term damage of the substance on the cells. The second is related to the short-term effects of alcohol overdose which manifests in what is commonly called,
the state of drunkenness and its consequences. The last mechanism is related to alcoholism and alcohol use disorder (World Health Organization 2014).

Ethanol, through its metabolites toxicity, notably acetaldehyde and acetic acid (Murty 2004), has a direct damaging effect on the DNA (Brooks 1997). Through this mechanism, the majority of the body organs get damaged with the long-term exposure to alcohol (Caan & Belleruche 2002). Brain cells are damaged through mechanisms including neuroinflammatory induced apoptosis (Pascual et al. 2007) causing brain development impairment, dementia and cognitive disorders (Panza et al. 2009). In cardiovascular system, the risk of heart attack increases with prolonged moderate to heavy drinking (Krenz & Korthuis 2012). Heavy drinking also enormously rises the risk of hemorrhagic strokes (Emberson & Bennett 2006). In addition, major public health issues worldwide such as liver cirrhosis, alcohol liver disease, and the risk of developing cancer are, in many cases, direct consequences of alcohol consumption (World Health Organization 2014).

The distribution of alcohol use and consumption worldwide varies widely from country to country from countries mostly abstinent such as Algeria and Morocco to regions where heavy drinking is the norm such as in some Ex-Soviet Union countries and a few other European countries. According to data from 2016, prevalence of current drinkers was the highest in high income countries (77%) and the lowest in low income countries (14%). As a general trend, men drank significantly more than women. The consumption gap between men and women was found to increase the poorer the country is (Griswold et al. 2018).

People drink alcohol for many different reasons. In countries where drinking alcohol is common, people mainly drink it either to cope with stress or because of social influences (Abbey et al. 1993).

In terms of disease burden, 4.5% of all global deaths, or roughly 2.8 million deaths, were attributable to the harmful use of alcohol in 2016 (Griswold et al. 2018) with a mortality rate of nearly 40 deaths per 100 000 in 2017 (GBD 2017 Causes of Death Collaborators 2018). The disability-adjusted life years lost to alcohol were estimated at around 3.8% of the global DALYs for 2016 (Griswold et al. 2018).
The proportion of mortality varied as well between regions. The mortality distribution pattern matched the map of the distribution of consumption (World Health Organization 2004). Former Soviet Union countries had the highest mortality rate attributable to alcohol consumption while, as expected, the eastern Mediterranean region, which had the lower consumption of alcohol worldwide, showed the lower rates of mortality (Griswold et al. 2018).

The relation between alcohol consumption distribution and mortality was known since 1926 as well as the importance of drinking patterns (Pearl 1926). Griswold et al. (2018) also underlined the contribution of patterns of drinking, especially heavy drinking occasions, in the disease burden. An article published by Alcohol Research & Health (Chartier & Caetano 2010) details the contrast of patterns of drinking between different ethnicities in the US and associates binge drinking with higher morbidity and mortality rates.

Despite the heavy burden that alcohol is weighing globally, as a public health issue, alcohol related consequences are reversible. Strengthening laws and policies has been proven to be potentially effective in limiting that burden and the major players in the alcohol battle seem to be the policy makers against the alcohol industry. The worldwide tolerance of alcohol is, if the cultural and historical circumstances are excluded, unjustifiable. And, while many suggest that light alcohol consumption carries some health benefits, as Griswold et al. (2018) put it “the safest level of drinking is none”.

2.3.3 Physical activity and other health behaviors

In addition to nutritional choices, smoking tobacco, and drinking alcohol, many other behavioral risk factors influence health. Physical activity is the next important lifestyle-related risk factor worth mentioning.

Physical activity determines many aspects of health and disease. While, it is mostly related to cardiovascular disease, it modulates other diseases such as diabetes, cancer, Alzheimer disease, etc. through obesity and other pathways (Lobelo et al. 2018).

The Global Burden of Disease study determined that more than 1.2 million deaths in 2017 were attributed to low physical activity, 90% of which had cardiovascular diseases as a death cause
The general trends of physical activity levels globally have been declining for the past few decades (Katzmarzyk & Mason 2009) substantially contributing to the current burden of diseases (Lee et al. 2012).

Physical activity can include leisure time physical activity (such as hiking, running, and swimming), occupational physical activity, and transportation-related physical activity. WHO recommends, in addition to two or three sessions of muscle-toning, 150 minutes of moderate-intensity or at least 75 minutes of vigorous-intensity aerobic physical activity every week for adults younger than 65 years of age (World Health Organization 2019). A study that followed over half a million American for a median of 10 years found that a moderate level of physical activity combined with a normal body mass index were associated with a gain in life expectancy of over 7 years in comparison to being physically inactive and obese (Moore et al. 2012).

As discussed previously in the chapter ‘obesity’, low levels of leisure time physical activity do not represent the only determinant of sedentary behavior. Television viewing, computer use, and reading, when done in excess, have been found to be associated to increased risk of obesity and chronic diseases (Hu 2003).

As to predicting life-expectancy, physical activity level does not only represent a behavioral status, but it also gives an overview of the body capacities to perform. The results of a new study suggest that adults who are able to complete more than 40 push-ups have significantly lower rate of CVD incidents than those who can only complete 10 push-ups or less (Yang et al. 2019). This double role of physical activity as a behavior and as an indicator of physical capacity might be source of confounding in epidemiological studies.

Malnutrition is another lifestyle-related risk factor, which, again - as previously mentioned, constitute a heavy burden on mortality ranking third in term of deaths attributed to lifestyle-related risks (Figure 9) (GBD 2017 Causes of Death Collaborators 2018).
Although the place of malnutrition is narrower in countries with high socio-development index, food insecurity is still prevalent in many developed countries. In 2017, about 12% of US households (40 million individuals) were food-insecure (Coleman-Jensen et al. 2018). The shortage often concerns quality of diet rather than quantity creating what can be referred to as ‘the dual burden of malnutrition’ where highly caloric that is deprived of essential nutrients dominates the diet causing a chronic disease – often obesity – coupled with malnutrition (Bowers et al. 2018).

Other lifestyle-related behaviors that affect public health in a significant way and represent a source of preventable premature mortality include drug use, which can cause violence, gastro-intestinal diseases, psychiatric disorders, and even some cancers. Unsafe sex also cause some burden in mortality mainly through HIV/AIDS in some low socio-economic populations and through cervical cancer and other STDs (GBD 2017 Causes of Death Collaborators 2018).

It is also worth mentioning that, even though many studies assessed the contribution of individual lifestyle-related behaviors on health and mortality, it seems that there are not many studies that
studied the combined effects of different behaviors on health and mortality (Fazel-Tabar Malekshah et al. 2016).

2.4 Midlife

Midlife, also called middle age, refers to the period of a person’s age from 45 years to 60 or 65 years.

Middle age is usually the period of life in which some of the most common chronic diseases start appearing (La Vecchia et al. 2011) as a result of the cumulative effects of childhood circumstances, young adult’s lifestyle, and, from a life-course perspective, the individual’s socio-environmental conditions (Fuller-Iglesias et al. 2009). As Ory et al (2014) had nicely put it citing from Satariano (2005): “from a health perspective, midlife represents a watershed, the period at which host immunity begins to decline and the effects of behavioral, social, and environmental risks for ill health begin to accumulate”.

In Finland, about a third of the population belongs to the midlife age group (Figure 10). This population group tend to be responsible of a big portion of health expenditures. In addition, due to the common changes in both their physiology and life circumstances, this group is more likely to adopt an unhealthy lifestyle (White et al. 2014).
It could be thought that it is possibly too late to attempt to affect behavior of middle-aged adults and target health promotion campaigns to this age category. However, evidence shows that interventions at midlife might considerably reduce premature death and potentially increase life expectancy. For instance, smoking cessation at midlife has been found to reduce most of the lung cancer risk attributed to cigarette smoking (Peto et al. 2000). Ory et al. (2014) saw that midlife is a great opportunity to make positive changes to health lifestyles since this age group would already have a ‘glimpse of later life’.

3 AIM OF THE STUDY

The purpose of this study is to evaluate the combined effect of the main health behaviors and to determine which ones are the most important predictors of life expectancy.
The secondary aim of the study is to develop a predictive model that permits the expression of the effects of health behaviors with life-expectancy as an outcome.

4 RESEARCH METHODOLOGY

4.1 Study design and settings

The Kuopio Ischaemic Heart Disease Risk Factor Cohort Study (KIHD)
The study is based on the prospective, population-based Kuopio Ischaemic Heart Disease Risk Factor Cohort Study (KIHD). The original purpose of the cohort was to explore the cardiovascular risk factors that were common in Kuopio city and its surroundings. Its study protocol was approved by the Research Ethics Committee of the University of Kuopio – today the University of Eastern Finland after merging with the University of Joensuu. Men aged between 42 to 60 years were randomly sampled from the region of Kuopio in Eastern Finland, and among the 3235 eligible individuals, 2682 men were enrolled to the cohort at some point between March 1984 and December 1989 (Salonen 1988). After baseline, the follow-up study continued collecting data on the health status of participants through reexaminations after 4, 11, and 20 years, and via its linkages with national covering registries including the Cause of Death registry maintained by Statistics Finland, the Finnish Cancer Registry, and the Care Register for Health Care regarding inpatient hospitalizations maintained by the National Institute for Health and Welfare. The registries served mainly in monitoring causes of deaths, cancer diagnoses, inpatient diagnoses, and hospital treatment episodes. Despite its original cardiovascular-oriented original purpose, the study turned into a multidisciplinary project involving a considerable range of data on a wide variety of health determinants (Kauhanen 2013). The cohort is still ongoing as of May 2019.

4.2 Measurement of independent factors

All baseline measurements were collected at the beginning of enrollment of each participant. The independent factors of interest for our study are represented by the following health-related behavioral factors: diet, tobacco smoking, alcohol drinking, and physical activity level. These four parameters were chosen because they are the main behavioral contributors to risk according to the literature (GBD 2017 Causes of Death Collaborators 2018).
Diet was assessed at baseline based on a self-reported questionnaire and a food record of four days. In practice, when the subjects came for blood sampling during the phase of baseline examinations, they were given instructions by a nutritionist to record their food intake for 4 days. The nutritionist checked then their records. The software NUTRICA (version 2.5), which was developed by National Public Health Institute, Turku – Finland, and included Finnish values of 30 nutrients composing foods, was used to quantitatively estimate averages of intake of different nutrients for each participant (Voutilainen et al. 2001). Baltic Sea Diet Score (BSDS) is used as the indicator of the healthiness of the dietary behavior in our study.

The self-reported questionnaire at baseline also included questions related to current and past health-related behaviors. To assess tobacco smoking for example, current and previous smoking status was reported, as well as the number of daily packets of cigarette smoked. The packet-year indicator was then calculated to quantitatively reflect the cumulative exposure to tobacco. In our study, since we have found that even low cigarette consumption can lead to significantly negative health outcomes (Hackshaw et al. 2018), we only consider the status of smoking at baseline as a proxy of tobacco smoking despite the daily quantity of cigarettes smoked.

Detailed self-reported questionnaires assessed current alcohol drinking status as well as the weekly quantity of alcohol usually consumed. In our study, alcohol consumption level is treated as a continuous variable and measured in units of 100 grams per week.

On the other hand, physical activity level was thoroughly assessed via different methods. The self-reported questionnaire included questions on the daily physical activities. A 12-month history, a 7-day recall, and a 24-hour self-recording of physical activity supported the data. In addition to that, an interview was conducted at baseline to help estimate the amount of physical activity involved in daily work. Direct estimates were then conducted on the participants who underwent some active exercises accompanied by respiratory gas measurements. These estimates were then found to correlate quite well with the self-reported data (Salonen & Lakka 1987). Yearly indicators of physical activity, notably total leisure time physical activity, conditioning leisure time physical activity, and their metabolic equivalent (MET) were then computed. The WHO-recommended 150 minutes of weekly moderate intensity leisure time physical activity (World Health Organization 2019) is roughly equivalent to 500 MET-minutes per week, and the weekly recommended 75
minutes of vigorous intensity physical activity is equivalent to 1000 MET-minutes per week. For our study, we are, thus, using yearly MET-hours as a continuous indicator of physical activity since it takes into consideration the intensity of physical activity in its calculation. 500 to 1000 MET-minutes per week corresponds to 434.5 to 869 MET-hours per year (Salonen & Lakka 1987).

### 4.3 Measurement of mortality and morbidity

Information on mortality was collected through a link to the Cause of Death Registry. Ascertainment of follow-up events was done using the Finnish equivalent of the social security number.

Morbidity data was assessed at baseline through the self-reported questionnaires and included information on present morbidities and previous health events. This information is used to compute the Charlson Comorbidity Index (CCI) (Quan et al. 2005) to represent the morbidity indicator in our study. We estimate that 2005 version of CCI is a better indicator of longer term mortality than the updated version of Quan et al. (2011) specifically adapted for 1-year in-hospital mortality.

After baseline, data on morbidities and health events is fed to the KIHD database through linkage to the Care Register for Healthcare regarding inpatient hospitalizations and the Finnish Cancer Registry. This diagnostic data, coded in the International Classification of Diseases (ICD) ICD-9, ICD-10, and the International Classification of Diseases for Oncology (ICD-O) ICD-O-3, is used in our study to compute a ‘follow-up CCI’ for the purpose of evaluating validity and reliability of the baseline CCI derived from the self-reported questionnaires (page 68).

The maximum duration of follow-up was close to 32 years and 10 months.

### 4.4 Sample size review

Our study is a cohort following healthy subjects with different rates of exposure to behavioral risk factors over a period of time exceeding 30 years in some cases. As the interest of our study is to evaluate changes in life expectancy – or mortality as a proxy - due to selected behavioral risk factors, a sample size that is sufficient enough to provide statistical results depends in its calculation on mortality rates due to these behavioral factors as reported from previous studies. When studying
the effects of multiple risk factors, the most harmful of these factors can be used for sample size calculation.

According to the literature (GBD 2017 Causes of Death Collaborators 2018), diet is the most important contributor behavior, among our selected behaviors, to mortality in Finland followed by tobacco smoking. Due to ease of calculation and data retrieval, we use previous tobacco smoking for sample size estimation.

We extract smoking-related mortality data from the 1951 prospective study on male British doctors (Doll et al. 2004) to provide estimates of previous rates of outcome among exposed and non-exposed individuals and retrieve the following rates (Table 1).

<table>
<thead>
<tr>
<th>Rates in smokers</th>
<th>Rates in non-smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality at 65 years of age</td>
<td>6%</td>
</tr>
<tr>
<td>Mortality at 70 years of age</td>
<td>19%</td>
</tr>
<tr>
<td>Mortality at 75 years of age</td>
<td>42%</td>
</tr>
</tbody>
</table>

Table 1 Probability of death in smokers and non-smokers from the 1951 British doctors’ study

We may use the formula below (Abrouk 2018) to estimate the required sample size in each group in the case that the number of exposed equals that of the non-exposed.

\[
n = \frac{2(z_{\alpha/2} + z_\beta)^2 p(1-p)}{(p_1 - p_0)^2}
\]

\(n\): minimum sample size of each group if number of exposed equals number of non-exposed
\(p_1\): mortality rate in the exposed population (from the literature)
\(p_0\): mortality rate in the non-exposed population (from the literature)
\(p = (p_1 + p_0)/2\)
\(Z_{\alpha/2}\): standard normal variate for level of significance
\(Z_{\beta}\): standard normal variate for power or type 2 error
For p-value = 0.05, $Z_{a/2} = 1.96$
For power = 80%, $Z_\beta = 0.842$

The result: $n = 750$

However, since the number of smokers differ from that of the non-smokers in our study, we will apply the formula below (Abrouk 2018) to determine the minimum required sample size in the exposed group $n_E$ while $k$ corresponds to the ratio between the number in the non-exposed group and the exposed group.

$$n_E = \frac{k + 1}{2k} n$$

In our cohort, smoking attitudes at baseline are distributed by age groups as follows (Table 2).

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Never</th>
<th>Previous</th>
<th>Current smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>42 - 47</td>
<td>119</td>
<td>93</td>
<td>122</td>
</tr>
<tr>
<td>47 - 52</td>
<td>124</td>
<td>120</td>
<td>114</td>
</tr>
<tr>
<td>52 - 57</td>
<td>517</td>
<td>572</td>
<td>503</td>
</tr>
<tr>
<td>57 - 62</td>
<td>101</td>
<td>170</td>
<td>127</td>
</tr>
</tbody>
</table>

Table 2 Smoking status by age group at baseline

After calculation, the minimum number of subjects needed in the smoking group shall be 590. The sample size of our study fulfills this requirement.

### 4.5 Validating baseline Charlson Comorbidity Index (CCI)

Determining whether an instrument (index, score, or other forms of indicators) is correctly measuring what is meant to measure is often part of separate studies that follow well-codified processes. Such studies generally need to test content, construct, and criterion-related validity, as well as to verify the instrument’s reliability.
However, CCI is already well validated as an instrument of measure of comorbidities and the only limitation that we have in computing the baseline CCI is that we are calculating it based on self-reported questionnaires through which participants communicated their diagnoses rather than through ICD codes from hospital registries. As mentioned earlier, hospital registries’ ICD-9 and ICD-10 codes were fed to the study participants’ data as diagnoses were coming up during follow-up. This information is used in our study to calculate (Quan et al. 2005) ‘follow-up CCI’ for each patient based on diagnoses established up until December 2016.

Since both baseline CCI and follow-up CCI do not follow a normal distribution, we use a non-parametric test to examine the correlation between the two variables. Using R Programming language (R Core Team 2019), we have tested the correlation between CCI and follow-up CCI using Spearman correlation test and found that there is a statistically significant (p-value < 2.2e-16) positive correlation (r = 0.17) between the two variables (Figure 11).

Although the two variables are but weakly correlated, since follow-up CCI is expected indeed to have different values than baseline CCI as comorbidities get diagnosed with follow-up time, the positive value of r and the significance of the test suggest that baseline CCI can be used as a valid indicator of baseline comorbidities.
Figure 11 Correlation between baseline CCI and follow-up CCI

4.6 Statistical analysis

All calculations are done using R programming language Version 3.5.3 (11/03/2019) nickname “Great Truth” (R Core Team 2019) through the GUI interface of RStudio Version 1.1.463 – © 2009-2018 RStudio, Inc.

4.6.1 Cox regression survival-analysis

Cox proportional-hazards model is a statistical regression-based model (Cox 1972) that is commonly used in health research to test the association between one or more predictors and survival time.

Cox regression survival-analysis is used to determine how health-related behaviors determine the outcome of all-cause death. Smoking status is used as a categorical variable while alcohol level and physical activity level are used as a continuous variable measured in unites of 100 grams per week and MET-hours per year respectively. BSDS is computed as an indicator of healthy diet and
is used as a continuous variable. The four previously mentioned variables constitute the independent variables of the model.

Covariates are represented by age - as a categorical variable (42 – 47 years, 47 – 52 years, 52 – 57 years, and 57 – 62 years), and CCI - as a continuous variable. These covariates were found to significantly correlate with all-cause mortality using Spearman’s correlation test. BMI is split into categories: normal weight [18.5 – 25], underweight [≤18.5], overweight1 [25 – 27.5], overweight2 [27.5 – 30] and obese [≥30].

There are two study participants with a baseline BMI that is slightly below normal. Since they are too few to provide statistical significance and because the literature review does not suggest that slightly below average BMI influences mortality, for the sake of better statistical results, we added them to normal weight category.

The R Survival Analysis package by Therneau T (2000) version 2.43-3 published on 26.11.2018 which includes Cox models, Kaplan-Meier, and Aalen-Johansen multi-state curves was used to build, diagnose, and analyze the Cox regression model.

The formula of the study’s main model is:
Surv(daysofsurv,status)~agecat+fit+CCI+smoking+alcohol+physical+BSDS,kihdsurv
where daysofsurv represents the time of survival, agecat represents age categories, and fit represents BMI categories.

From the 2682 study participants, 66 were not included in the analysis due to missing values. Final number of participants analyzed n=2616. The mean follow-up time was of 23.3 years. During the total length of follow-up of nearly 32 years and 10 months, 1479 deaths were recorded.

4.6.2 Model diagnostics

Cox regression survival models, although do not assume any specific survival model, are not truly non-parametric models. Some conditions need to be verified before assuming that the Cox regression model properly describes the data it is fitted to.
The main assumption for Cox models is the proportional hazards (PH) assumption. It assumes that the effects of the covariates and independent variables on mortality are not varying over time. In theory, this assumption is not met in our study because many of our predictive variables are not constant over time – the case of most predictors in clinical research (Zhang et al. 2018). A study participant who is a smoker at baseline will not necessarily remain a smoker or smoke the same amount along the time of follow-up (Pinsky et al. 2015). However, our interest is time to mortality and life-expectancy prediction based on baseline parameters, and thus, in the scope of our study, we are using Cox regression survival modelling with the assumption that baseline parameters would remain constant over time. An extension of our study could benefit from data on changes of behaviors and covariates over time and use time-dependent Cox models to better the prediction (Therneau & Grambsch 2000, Zhang et al. 2018).

Schoenfeld method is used to examine the proportional hazard assumption. Graphs representing scaled Schoenfeld residuals for each of the model’s predictors through time have been generated for analysis accompanied by Schoenfeld tests of individual covariates as well as the model in global. The assumption of proportional hazards was verified for all the covariates except smoking and CCI. Schoenfeld test for the global model has also failed to verify the assumption of proportional hazards. The model is, thus, not totally meeting the assumptions as it is but might benefit from stratification by smoking status and some tuning.

5 RESULTS

The following table (Table 3), in addition to the number of events (deaths) in each group, summarizes the main baseline characteristics of the study population included in the analysis. Except for age categories distribution, baseline characteristics differed significantly between smokers and non-smokers.

Table 3 Baseline characteristics of the study population stratified by smoking status

<table>
<thead>
<tr>
<th></th>
<th>Non-smokers</th>
<th>Smokers</th>
<th>Total</th>
<th>P-values b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>1784</td>
<td>832</td>
<td>2616</td>
<td></td>
</tr>
<tr>
<td>Number of events (%)</td>
<td>890 (49.9)</td>
<td>589 (70.8)</td>
<td>1479</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Median age at baseline belonged to the age category 52-57 and age distribution in general did not differ much between smoking groups at baseline (p-value = 0.302). The smoker group had a significantly (p-value < 0.001) lower proportion of individuals with increased BMI (60.5%) than non-smokers group (73.8%). Moreover, in comparison to non-smokers, smokers tended to exercise less (p-value = 0.003), eat less healthy (p-value < 0.001), consume double the amount of alcohol (p-value < 0.001), and have more morbidities (p-value = 0.019) at baseline.

### 5.1 Main model

The main model, as mentioned before, is a multivariate Cox regression model accounting for age categories, smoking status, alcohol consumption, BSDS, BMI categories, and CCI. Analysis of the model showed statistically significant effect of smoking, alcohol drinking, and diet on time to mortality (Table 4). Obesity and initial morbidities were also found associated with higher mortality. The model has a statistically significant (p-value ≤ 2x10^{-16}) Wald test of 617.7 (and Likelihood ratio test of 612.6). R-squared in Cox regression might not measure the goodness of fit.
the way it does in linear regression for example (Schemper & Henderson 2000), but it is worth mentioning that our model’s R-square = 0.209 which may suggest that the model explains 20.9% of the variation in time to mortality.

Table 4 Cox Proportional Hazards main model

<table>
<thead>
<tr>
<th>Age category</th>
<th>Hazard Ratio (HR)</th>
<th>95% Confidence Intervals</th>
<th>P-values $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>42-47 a</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>47-52</td>
<td>1.62</td>
<td>1.22 – 2.17</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>52-57</td>
<td>3.24</td>
<td>2.57 – 4.10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>57-62</td>
<td>5.87</td>
<td>4.55 – 7.57</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BMI category</th>
<th>Hazard Ratio (HR)</th>
<th>95% Confidence Intervals</th>
<th>P-values $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal weight [≤25] a</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>overweight 1 ]25 - 27.5]</td>
<td>0.93</td>
<td>0.82 – 1.07</td>
<td>0.304</td>
</tr>
<tr>
<td>overweight 2 ]27.5 – 30[</td>
<td>1.16</td>
<td>1.00 – 1.34</td>
<td>0.054</td>
</tr>
<tr>
<td>obese [≥30]</td>
<td>1.40</td>
<td>1.20 – 1.63</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Hazard Ratio (HR)</th>
<th>95% Confidence Intervals</th>
<th>P-values $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.91</td>
<td>1.71 – 2.13</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical activity (MET hour/day) $^c$</th>
<th>Hazard Ratio (HR)</th>
<th>95% Confidence Intervals</th>
<th>P-values $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSDS</td>
<td>0.97</td>
<td>0.96 – 0.98</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CCI</td>
<td>1.14</td>
<td>1.10 – 1.19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Alcohol consumption (100g/week)</td>
<td>1.12</td>
<td>1.085 – 1.16</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

$^a$ reference category for hazard ratio estimation

$^b$ p-values of the Z-tests (Wald statistics) related to each covariate

$^c$ The unit MET hour/year was changed to MET hour/day in order to more appropriately show the effect size

Further stratification by age category was done and the resulting effect sizes of the analysis expressed in hazard ratios are illustrated as a Forest plot (Figure 12) without showing much difference in the results.
Figure 12 Forest plot illustrating Hazard ratios of the main model's covariates (alc100gweek corresponds to Alcohol level measured in units of 100g per week & physday corresponds to Physical activity level measured in MET-hours per day)

The model’s concordance index of 0.687 (0.64 with age-stratification) indicates a high predictive accuracy of outcome.

The analysis showed a significant association (p < 0.001) between smoking and lower survival with a HR of 1.91 (95% CI 1.71 – 2.13). Study participants with high BSDS – reflector of a healthy diet – had, with statistical significance (p < 0.001), better survival rates than participants with low values of BSDS (HR 0.97, 95% CI 0.96 – 0.98). Alcohol consumption (units of 100 grams per week) was also associated with lower survival with a HR of 1.12 (95% CI 1.08 – 1.15) with a high statistical significance (p < 0.001).
The model’s covariates (Age categories, BMI categories, and CCI) were also associated with a statistically significant influence on time to mortality. CCI for example, was linked with a statistically significant 14% increase in mortality with each unit of the index (HR 1.15, p-value < 0.001). However, a BMI from 25 to 30 corresponding to BMI categories overweight 1 and overweight 2 did not show enough significance in outcome prediction in our main model (p-values of 0.304 and 0.054 respectively).

Similarly, physical activity level – measured in metabolic hours per year (or day) as a continuous variable – was unable to demonstrate significant association with time to mortality (p-value = 0.365).

The following graphs (Figure 13) represent Kaplan-Meier survival plots illustrating the changes in survival probability of the analyzed population along the follow-up time as described by the main Cox regression model.

![Figure 13 Main model's Kaplan-Meier survival curve.](image)

The bottom graph shows a stratification by smoking status. 0: nonsmokers. 1: smokers.

Time is estimated in years of follow-up.
Aalen additive models as illustrated on Figure 14 describe the effect of different categories of variables on the probability of death over the follow-up time. The deleterious effect of smoking and obesity seems to significantly increase and cumulate over time.

Figure 14 Aalen regression plots illustrating how smoking status, age categories, and BMI categories influence survival

In order to further explore the effect of physical activity on time to mortality, and as an attempt to obtain more statistical significance, we have broken the continuous physical activity variable into 8 levels of yearly MET-hours of physical activity. We modified our Cox proportional hazards model to include these levels instead. The modification slightly improved R-squared of the model from 0.209 to 0.212 but physical activity levels remained statistically non-significant except in some strata. We report here (Figure 15) the Forest plot of this new model in the stratum of smokers aged 52 to 57. The protective effect of physical activity shows significant beneficial effect on survival at levels 1650-2250 MET hours per year (HRs = 0.57 p-value < 0.01) in reference to very low level of physical activity (less than 270 metabolic hours per year).
Figure 15 Forest plot of the main Cox regression model with physical activity broken into levels. Stratum: smokers aged 52-57

The tables below (Table 5) illustrate the changes of risk of death after 20 years of follow-up through changes in Risk Score, an index that we derived from exp(lp), modulated by changes in health behaviors with selected conditions at CCI = 0 since this value is the median for baseline. The scores are generated by prediction of survival in generated synthetic cases using the main Cox regression model with physical activity broken into levels.
Table 5 Predicted Risk Score for multiple behavioral risk factors. Green: score in favor of survival. Red: score in favor of mortality.

Score = rounded exp(lp) risk score at 20 years of follow up multiplied by 10
The following values concern age category 47-52 at Alcohol level = 0

<table>
<thead>
<tr>
<th>Physical activity level (MET-hours per year)</th>
<th>normal weight nonsmoker</th>
<th>smoker</th>
<th>overweight1 nonsmoker</th>
<th>smoker</th>
<th>overweight2 nonsmoker</th>
<th>smoker</th>
<th>obese nonsmoker</th>
<th>smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>150 BSDS = 5</td>
<td>4.9</td>
<td>9.5</td>
<td>4.6</td>
<td>8.8</td>
<td>5.7</td>
<td>10.9</td>
<td>7.0</td>
<td>13.3</td>
</tr>
<tr>
<td>350 BSDS = 5</td>
<td>4.1</td>
<td>7.9</td>
<td>3.8</td>
<td>7.3</td>
<td>4.7</td>
<td>9.1</td>
<td>5.8</td>
<td>11.1</td>
</tr>
<tr>
<td>650 BSDS = 5</td>
<td>4.6</td>
<td>8.8</td>
<td>4.3</td>
<td>8.2</td>
<td>5.3</td>
<td>10.1</td>
<td>6.5</td>
<td>12.4</td>
</tr>
<tr>
<td>1000 BSDS = 5</td>
<td>4.4</td>
<td>8.4</td>
<td>4.0</td>
<td>7.8</td>
<td>5.0</td>
<td>9.6</td>
<td>6.1</td>
<td>11.8</td>
</tr>
<tr>
<td>1800 BSDS = 5</td>
<td>4.0</td>
<td>7.7</td>
<td>3.7</td>
<td>7.2</td>
<td>4.6</td>
<td>8.9</td>
<td>5.7</td>
<td>10.8</td>
</tr>
<tr>
<td>3000 BSDS = 5</td>
<td>4.2</td>
<td>8.1</td>
<td>3.9</td>
<td>7.5</td>
<td>4.9</td>
<td>9.3</td>
<td>5.9</td>
<td>11.4</td>
</tr>
<tr>
<td>150 BSDS = 10</td>
<td>4.3</td>
<td>8.3</td>
<td>4.0</td>
<td>7.7</td>
<td>5.0</td>
<td>9.5</td>
<td>6.1</td>
<td>11.6</td>
</tr>
<tr>
<td>350 BSDS = 10</td>
<td>3.6</td>
<td>6.9</td>
<td>3.3</td>
<td>6.4</td>
<td>4.1</td>
<td>7.9</td>
<td>5.0</td>
<td>9.6</td>
</tr>
<tr>
<td>650 BSDS = 10</td>
<td>4.0</td>
<td>7.7</td>
<td>3.7</td>
<td>7.1</td>
<td>4.6</td>
<td>8.8</td>
<td>5.6</td>
<td>10.8</td>
</tr>
<tr>
<td>1000 BSDS = 10</td>
<td>3.8</td>
<td>7.3</td>
<td>3.5</td>
<td>6.8</td>
<td>4.4</td>
<td>8.4</td>
<td>5.3</td>
<td>10.2</td>
</tr>
<tr>
<td>1800 BSDS = 10</td>
<td>3.5</td>
<td>6.7</td>
<td>3.3</td>
<td>6.2</td>
<td>4.0</td>
<td>7.7</td>
<td>4.9</td>
<td>9.4</td>
</tr>
<tr>
<td>3000 BSDS = 10</td>
<td>3.7</td>
<td>7.0</td>
<td>3.4</td>
<td>6.5</td>
<td>4.2</td>
<td>8.1</td>
<td>5.2</td>
<td>9.9</td>
</tr>
<tr>
<td>150 BSDS = 15</td>
<td>3.8</td>
<td>7.2</td>
<td>3.5</td>
<td>6.7</td>
<td>4.3</td>
<td>8.3</td>
<td>5.3</td>
<td>10.1</td>
</tr>
<tr>
<td>350 BSDS = 15</td>
<td>3.1</td>
<td>6.0</td>
<td>2.9</td>
<td>5.5</td>
<td>3.6</td>
<td>6.9</td>
<td>4.4</td>
<td>8.4</td>
</tr>
<tr>
<td>650 BSDS = 15</td>
<td>3.5</td>
<td>6.7</td>
<td>3.2</td>
<td>6.2</td>
<td>4.0</td>
<td>7.7</td>
<td>4.9</td>
<td>9.4</td>
</tr>
<tr>
<td>1000 BSDS = 15</td>
<td>3.3</td>
<td>6.3</td>
<td>3.1</td>
<td>5.9</td>
<td>3.8</td>
<td>7.3</td>
<td>4.6</td>
<td>8.9</td>
</tr>
<tr>
<td>1800 BSDS = 15</td>
<td>3.1</td>
<td>5.8</td>
<td>2.8</td>
<td>5.4</td>
<td>3.5</td>
<td>6.7</td>
<td>4.3</td>
<td>8.2</td>
</tr>
<tr>
<td>3000 BSDS = 15</td>
<td>3.2</td>
<td>6.1</td>
<td>3.0</td>
<td>5.7</td>
<td>3.7</td>
<td>7.1</td>
<td>4.5</td>
<td>8.6</td>
</tr>
<tr>
<td>150 BSDS = 20</td>
<td>3.3</td>
<td>6.3</td>
<td>3.0</td>
<td>5.8</td>
<td>3.8</td>
<td>7.2</td>
<td>4.6</td>
<td>8.8</td>
</tr>
<tr>
<td>350 BSDS = 20</td>
<td>2.7</td>
<td>5.2</td>
<td>2.5</td>
<td>4.8</td>
<td>3.1</td>
<td>6.0</td>
<td>3.8</td>
<td>7.3</td>
</tr>
<tr>
<td>650 BSDS = 20</td>
<td>3.0</td>
<td>5.8</td>
<td>2.8</td>
<td>5.4</td>
<td>3.5</td>
<td>6.7</td>
<td>4.3</td>
<td>8.2</td>
</tr>
<tr>
<td>1000 BSDS = 20</td>
<td>2.9</td>
<td>5.5</td>
<td>2.7</td>
<td>5.1</td>
<td>3.3</td>
<td>6.4</td>
<td>4.0</td>
<td>7.8</td>
</tr>
<tr>
<td>1800 BSDS = 20</td>
<td>2.7</td>
<td>5.1</td>
<td>2.5</td>
<td>4.7</td>
<td>3.1</td>
<td>5.9</td>
<td>3.7</td>
<td>7.2</td>
</tr>
<tr>
<td>3000 BSDS = 20</td>
<td>2.8</td>
<td>5.3</td>
<td>2.6</td>
<td>5.0</td>
<td>3.2</td>
<td>6.2</td>
<td>3.9</td>
<td>7.5</td>
</tr>
</tbody>
</table>
Score = rounded \(\exp(lp)\) risk score at 20 years of follow up multiplied by 10
The following values concern age category 42-47 at Physical activity level = 650

<table>
<thead>
<tr>
<th>Alcohol (100g/week)</th>
<th>normal weight</th>
<th>overweight1</th>
<th>overweight2</th>
<th>obese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nonsmoker</td>
<td>smoker</td>
<td>nonsmoker</td>
<td>smoker</td>
</tr>
<tr>
<td><strong>BSDS = 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2.8</td>
<td>5.4</td>
<td>2.6</td>
<td>5.0</td>
</tr>
<tr>
<td>0.16</td>
<td>2.9</td>
<td>5.5</td>
<td>2.7</td>
<td>5.1</td>
</tr>
<tr>
<td>1</td>
<td>3.1</td>
<td>6.0</td>
<td>2.9</td>
<td>5.6</td>
</tr>
<tr>
<td>3</td>
<td>3.9</td>
<td>7.6</td>
<td>3.7</td>
<td>7.0</td>
</tr>
<tr>
<td><strong>BSDS = 10</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2.4</td>
<td>4.7</td>
<td>2.3</td>
<td>4.4</td>
</tr>
<tr>
<td>0.16</td>
<td>2.5</td>
<td>4.8</td>
<td>2.3</td>
<td>4.4</td>
</tr>
<tr>
<td>1</td>
<td>2.7</td>
<td>5.2</td>
<td>2.5</td>
<td>4.9</td>
</tr>
<tr>
<td>3</td>
<td>3.4</td>
<td>6.6</td>
<td>3.2</td>
<td>6.1</td>
</tr>
<tr>
<td><strong>BSDS = 15</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2.1</td>
<td>4.1</td>
<td>2.0</td>
<td>3.8</td>
</tr>
<tr>
<td>0.16</td>
<td>2.2</td>
<td>4.2</td>
<td>2.0</td>
<td>3.9</td>
</tr>
<tr>
<td>1</td>
<td>2.4</td>
<td>4.6</td>
<td>2.2</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>5.7</td>
<td>2.8</td>
<td>5.3</td>
</tr>
<tr>
<td><strong>BSDS = 20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.9</td>
<td>3.6</td>
<td>1.7</td>
<td>3.3</td>
</tr>
<tr>
<td>0.16</td>
<td>1.9</td>
<td>3.6</td>
<td>1.8</td>
<td>3.4</td>
</tr>
<tr>
<td>1</td>
<td>2.1</td>
<td>4.0</td>
<td>1.9</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>2.6</td>
<td>5.0</td>
<td>2.4</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Score = rounded \(\exp(lp)\) risk score at 20 years of follow up multiplied by 10
The following values concern age category 47-52 at Physical activity level = 650

<table>
<thead>
<tr>
<th>Alcohol (100g/week)</th>
<th>normal weight</th>
<th>overweight1</th>
<th>overweight2</th>
<th>obese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nonsmoker</td>
<td>smoker</td>
<td>nonsmoker</td>
<td>smoker</td>
</tr>
<tr>
<td><strong>BSDS = 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4.6</td>
<td>8.8</td>
<td>4.3</td>
<td>8.2</td>
</tr>
<tr>
<td>0.16</td>
<td>4.7</td>
<td>9.0</td>
<td>4.3</td>
<td>8.3</td>
</tr>
<tr>
<td>1</td>
<td>5.1</td>
<td>9.9</td>
<td>4.8</td>
<td>9.2</td>
</tr>
<tr>
<td>3</td>
<td>6.5</td>
<td>12.4</td>
<td>6.0</td>
<td>11.5</td>
</tr>
<tr>
<td><strong>BSDS = 10</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4.0</td>
<td>7.7</td>
<td>3.7</td>
<td>7.1</td>
</tr>
<tr>
<td>0.16</td>
<td>4.1</td>
<td>7.8</td>
<td>3.8</td>
<td>7.3</td>
</tr>
<tr>
<td>1</td>
<td>4.5</td>
<td>8.6</td>
<td>4.2</td>
<td>8.0</td>
</tr>
<tr>
<td>3</td>
<td>5.6</td>
<td>10.8</td>
<td>5.2</td>
<td>10.0</td>
</tr>
<tr>
<td><strong>BSDS = 15</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3.5</td>
<td>6.7</td>
<td>3.2</td>
<td>6.2</td>
</tr>
<tr>
<td>0.16</td>
<td>3.5</td>
<td>6.8</td>
<td>3.3</td>
<td>6.3</td>
</tr>
<tr>
<td>1</td>
<td>3.9</td>
<td>7.5</td>
<td>3.6</td>
<td>6.9</td>
</tr>
<tr>
<td>3</td>
<td>4.9</td>
<td>9.4</td>
<td>4.6</td>
<td>8.7</td>
</tr>
<tr>
<td><strong>BSDS = 20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3.0</td>
<td>5.8</td>
<td>2.8</td>
<td>5.4</td>
</tr>
<tr>
<td>0.16</td>
<td>3.1</td>
<td>5.9</td>
<td>2.9</td>
<td>5.5</td>
</tr>
<tr>
<td>1</td>
<td>3.4</td>
<td>6.5</td>
<td>3.2</td>
<td>6.1</td>
</tr>
<tr>
<td>3</td>
<td>4.3</td>
<td>8.2</td>
<td>4.0</td>
<td>7.6</td>
</tr>
</tbody>
</table>
Throughout the tables, there is a clear trend of risk decrease with the increase of BSDS – equivalent to a better diet, and a clear increase of risk (nearly two folds) with smoking. While obesity and overweight2 are associated with an increased risk of mortality in comparison to normal weight, overweight1 seems to be paradoxically associated with a better survival in comparison to normal weight as a trend throughout the tables - previous analysis showed low statistical significance.
Regarding physical activity, it seems that there is a decrease of risk score with physical activity levels > 270 MET-hours per year in comparison to the reference level (<270 MET-hours per year) in favor of survival. Levels beyond 270, however, do not seem to be linearly correlated with survival and show irregular variation. For this reason, we displayed only one prediction table portraying physical activity.

On the other hand, increase in alcohol consumption was found to be associated with increase in mortality risk as shown on the four last figures of Table 5. A subject with an alcohol consumption of 300 grams/week is predicted to have an excess risk of 40% to die after 20 years in comparison to abstinent subjects.

![Forest plot of the main Cox regression model with BSDS broken into levels](image-url)
The continuous variable BSDS was also cut into levels and the analysis of the related Cox regression model generated the following Forest plot (Figure 16). Subjects with the healthiest dietary habits (BSDS between 20 and 25) were found to have a significantly (p-value 0.027) lower mortality risk (HR 0.61 CI 0.39 – 0.94) in comparison to the reference category with the less healthy dietary habits (BSDS between 0 and 5).

As to determine the life expectancy lost to unhealthy behavior, we have generated a dataset of two distinct attitudes toward health behaviors and used – again – prediction calculation based on the main model to estimate survival time.

The two distinct attitudes toward health behaviors were defined as follows. Healthy attitude: normal weight, nonsmoker, BSDS = 22, abstainer from alcohol. Unhealthy attitude: obese, smoker, BSDS = 3, alcohol = 500g/week. Age category 52-57, CCI = 0, and physical activity = 600 MET/hour per year are used for both groups. The following graph (Figure 17) is a plot of changes in survival probability throughout follow-up time.

![Predicted survival in ideal Healthy and Unhealthy Individuals](image)

Figure 17 Predicted survival difference between ideal healthy and most unhealthy individuals from a generated dataset
Figure 18, on the other hand, plots the changes in survival probability throughout follow-up time of another similarly generated dataset but in which the unhealthy group is set to the same BMI category (normal weight) as the healthy group.

![Graph](image)

Figure 18 Predicted survival difference between ideal healthy and normal weight most unhealthy individuals from a generated dataset

The graphs illustrate a gap of 17 to 20 years in predicted life-expectancy between the healthiest and the unhealthiest of generated cases at level of 50% of survival probability and a gap of 15 to 17 years in predicted life-expectancy at level of 80% survival probability.

Predictions were also made to compare two groups of synthetic (generated) subjects. Both groups are aged 52-57 with CCI = 0, normal weight, and optimal health behaviors (non-smokers, abstainers from alcohol, physical activity level at 600 MET-hours per year, BSDS = 25) but the second group has one health behavior changed into the unhealthy value of the third quartile of the study population. The resulting survival curves comparing the two groups are shown below (Figure 19). The choice of the value at the third quartile is meant to help in the comparability between factors of different natures.
Figure 19 Predicted survival difference between the ideal healthy and their peers who have one individual unhealthy behavior – predictions generated from a generated dataset

5.2 Smoking-stratified model

As recommended on model diagnosis chapter, we stratified by smoking status and dropped CCI from the model.

Kaplan-Meier survival plots on Figure 20, directly generated from the study population, show that, in all age categories, non-smokers tend to have a higher probability to live longer along the quasi-totality of the duration of follow-up with a difference of life expectancy averaging 8 years at 80% survival probability in favor of nonsmokers (Erreur ! Source du renvoi introuvable.).
Figure 20 Probability of survival by smoking status in different age categories with difference in life expectancy

Analysis of the Cox proportional hazards yielded the following results (Table 6):

**Table 6 Cox Proportional Hazards smoking-stratified model**

<table>
<thead>
<tr>
<th></th>
<th>Hazard Ratios in non-smokers</th>
<th>Hazard Ratios in smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(95% CI)</td>
<td>(95% CI)</td>
</tr>
<tr>
<td>Number of subjects</td>
<td>n = 1784</td>
<td>n = 832</td>
</tr>
<tr>
<td>Number of events</td>
<td>890</td>
<td>589</td>
</tr>
<tr>
<td>Concordance</td>
<td>0.67</td>
<td>0.64</td>
</tr>
<tr>
<td>R-square</td>
<td>0.157</td>
<td>0.158</td>
</tr>
<tr>
<td>Age category</td>
<td>42-47</td>
<td>47-52</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>BMI category</td>
<td>normal weight [≤25]</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>overweight 1 [25 - 27.5]</td>
<td>1.10 (0.91 – 1.33) b</td>
</tr>
<tr>
<td></td>
<td>overweight 2 [27.5 – 30]</td>
<td>1.46 (1.20 – 1.78) **</td>
</tr>
<tr>
<td></td>
<td>obese [≥30]</td>
<td>1.84 (1.51 – 2.24) **</td>
</tr>
<tr>
<td>Physical activity (MET hour/day) c</td>
<td>1.007 (0.99 – 1.022) b</td>
<td>1.009 (0.99 – 1.03) b</td>
</tr>
<tr>
<td>BSDS</td>
<td>0.99 (0.97 – 1.00) b</td>
<td>0.96 (0.94 – 0.98) **</td>
</tr>
<tr>
<td>Alcohol consumption (100g/week)</td>
<td>1.17 (1.10 – 1.23) **</td>
<td>1.10 (1.06 – 1.15) **</td>
</tr>
</tbody>
</table>

* p-value of the Z-test (Wald statistics) < 0.05
** p-value of the Z-test (Wald statistics) < 0.01
b p-value of the Z-test (Wald statistics) > 0.1

In a similar fashion as the main model, results from Cox proportional hazards of the smoking-stratified model show that advanced age and BMI categories are associated with higher rates of mortality than the reference categories in both the smoking and nonsmoking groups. Moreover, better diet was found to be significantly associated with higher survival in the smoking group (HR = 0.96, p-value < 0.01), but the results are not statistically significant in the nonsmoking group.

High alcohol consumption was found to be strongly associated (p-value < 0.01) with lower survival in both the nonsmoking and the smoking groups (HR 1.17 and 1.10 respectively).

Like the results from Table 5, The BMI category Overweight 1 (BMI from 25 to 27.5) is found associated with higher survival in smokers, but with a p-value slightly above the significance cutline (p-value = 0.078). Hazard ratios of BMI categories in the smoking group were in general found to be associated with lower risk than in the nonsmoking group.

The following Forest plots (Figure 21) represent the analysis results of the smoking-stratified model furtherly divided by age categories.
Figure 21 Survival Forest plots of the smoking-stratified model by age category
(alc100gweek corresponds to Alcohol level measured in units of 100g per week & physday corresponds to Physical activity level measured in MET-hours per day)
The double stratification by smoking status and age categories resulted in a loss of statistical significance due to the reduction of sample size.

However, some hazard ratios showed high statistical significance. Healthy diet for example was found to be associated with significant higher survival in smokers aged 52 to 57 and 57 to 62 (Hazard ratios of 0.96 and 0.95 respectively). The three higher than normal BMI categories were found to be associated with higher mortality in nonsmokers aged 52 to 57 (HR 1.33 for overweight 1, 1.77 for overweight 2, and 2.16 for obese. P-value in the three cases < 0.05). Paradoxically, slightly high BMI (overweight 1) was found to be associated with lower mortality in smokers aged 52 to 57 (HR = 0.72, p-value = 0.009) and in nonsmokers aged 57 to 62 but with lower significance in this case (HR = 0.70, p-value = 0.079).

Alcohol consumption yielded statistically significant Wald statistics tests in ¾ age categories of nonsmokers and ¼ categories of smokers showing a consistent increase of mortality risk with higher alcohol intake (HR from 1.12 to 1.51 for every 100 g of alcohol per week). Physical activity on the other hand was not found to have a statistically significant association with the outcome (p-value > 0.1).

6 DISCUSSION

6.1 Methodology, findings, and limitations

Our analysis of more than 2600 middle-aged men provided an insight of the relationship between the combined effects of selected behaviors and life-expectancy at midlife suggesting that behavioral risk factors may be responsible for a considerable gap in life-expectancy.

We examined health behaviors as both categorical and continuous variables. On categorical variables, we adopted simple categorization, such as dichotomizing in smoking behavior, as to simplify future reproducibility and give clearer meaning to findings, but we preferred continuous measures of exposure as to improve predictive performance (Altman & Royston 2006).

Cox regression analysis was used to examine the effects of health behaviors, in addition to age-category and selected covariates, on time-to-death as an outcome throughout the duration of follow-
up which reached nearly 33 years in some study participants with a mean follow-up time exceeding 23 years. We justified the selection of health behaviors and covariates through an extensive literature review, and then through correlation analysis for some factors.

One of the main findings that the study suggests is that the life-expectancy gap between healthy and unhealthy behaviors might reach 17 to 20 years of life lost (main model prediction on generated data with level of survival probability set to 50%). Similar to the findings by Manuel et al. (2016), smoking was found associated with a high hazard ratio among the health behaviors nearly doubling the risk of mortality (HR=1.91 95% CI 1.71 – 2.13). The years of mortality lost to smoking can reach 8 years as the Kaplan-Meier hazard curves of the main model of our study suggested.

The habit of smoking often begins in early adult life, but the frequency and length of usage differ from individual to another. Smoking status at baseline would, thus, not be reflective of the past habits of smoking nor of its future. Although we have adopted a simple smoker/nonsmoker categorization, the baseline questionnaire of our study and some follow-up examinations have collected data on the changes of smoking habits in the study participants providing a better capture of the cumulative exposure to smoking which can possibly be used with the same database in order to improve prediction accuracy and better assess the effect of smoking on life expectancy. As previously mentioned, Cox proportional models assume that exposure remains constant in amount and effect over time. Including changes in exposure might require a different approach that allows the inclusion of time-varying covariates in the regression model.

Despite that, the main model – as previously discussed – was found to potentially violate some of the Cox regression analysis assumptions. Stratification by smoking status was suggested on the section of model diagnosis as to attempt to correct assumption violation. Stratification by smoking status does not only compromise on the significance of the tests, but it also prevents us from properly draw conclusions on the effects of smoking on survival. However, the semi-parametric nature of the analysis allows some room for leaning over the robustness of the method, and thus, most of the study results are drawn from the analysis of the main model.

Diet on the other hand, topping the list of influencers of mortality attributed to health behaviors in Finland according to the Global Burden of Disease study (GBD 2017 Causes of Death...
Collaborators 2018), neared in its effects the influence of smoking on mortality in our study. With a BSDS hazard ratio of 0.96 to 0.97, the result reflects that 1 unit of improvement in BSDS allows a mortality risk reduction of 3 to 4%. Breaking BSDS by levels has shown that mid to high quality diet (BSDS > 10) was associated with a mortality risk reduction of up to 43% in comparison to very low quality diet (BSDS<5) (HR = 0.57, p-value < 0.001 for BSDS from 10 to 15) (Figure 16). However, with less than 4% of subjects having a very low-quality diet (BSDS<5), the study might lack proper apportionment to draw precise conclusions on the contrasts between the diet categories.

Although the adopted dietary assessment methods are trusted to reflect the dietary habits around the period of assessment, diets are also subject to changes in quantity, frequency, and quality over time, and calculating more BSDSs in more time points over the follow-up time – and including them in the analysis as time-varying covariates – might improve the model and its prediction accuracy.

Alcohol was found to be another statistically significant influencer of time to mortality with a risk increase of 12% for each unit of 100 grams of alcohol consumed weekly. However, with only 13% abstainers, the evaluation of the effects of alcohol consumption might also limit proper comparison between the subjects who consume alcohol and who do not. In addition, in a population where drinking is the norm, to abstain might be associated with a strong reason of abstention - a reason that might as well be associated with the outcome - opening a big parenthesis to confounding; Participants who have reported being abstainers at baseline for example might have been heavy drinkers before that and they have had to abstain as a measure against a morbid alcohol drinking abuse.

Also, as in dietary behavior, alcohol drinking behavior has patterns that differ in their influence of mortality. Binge drinking behavior, for example, is a pattern of alcohol abuse that was found to substantially increase mortality risk through alcohol poisoning, accidental injuries, pancreatitis, and other hazards (Courtney & Polich 2009). Counting on the average weekly consumed quantity only might not adequately reflect the effects of individual exposure to alcohol. Moreover, drinking habits tend to also change over time, probably toward the decrease with advance in ages as suggested by Molander et al. (2010) creating an overestimation of the cumulative alcohol exposure
in younger subjects in comparison to the older ones through follow-up time. Including patterns of drinking and treating the alcohol variable, again, as a time-varying factor might improve the model.

Physical activity, on the other hand, did not show significance in the main model as a continuous variable. But after breaking it into several levels, some levels have shown a statistically significant healthy effect on survival in some strata. Levels of physical activity going beyond 1350 MET hours per year - which were found significantly associated with beneficial health effects - exceed by far the WHO weekly recommended leisure-time physical activity to rates that are considered athletic. If not for the considerable loss of significance, the considerable effect size in comparison to the reference level makes physical activity as important factor as smoking, diet, and alcohol in terms of influence on time to mortality in our study.

A possible explanation for the loss of significance of the tests evaluating the effect size of physical activity on mortality is the small size of the reference category. In fact, less than 5% of the study participants had a physical activity level < 250 MET-hours per year at baseline. With more than 85% of the subjects reporting rates of leisure-time physical activity that satisfy the WHO recommendations, the studied population might be too homogeneous in term of physical activity to allow statistically significant comparisons. In addition, a good part of the beneficial effect of physical activity, and we mainly mean here weight loss, might already translate into a healthy BMI.

Moreover, considering that many of the study subjects were doing physically heavy occupations such as farming, a considerable proportion of physical activity might be lost in measurement if we content with leisure-time physical activity as a sole indicator of physical activity. However, occupational physical activity, which is characterized by physical efforts of lower intensity but constrained and long in duration in comparison to the physical efforts of leisure-time physical activity, was found to be paradoxically associated with a higher risk of morbimortality in comparison to leisure-time physical activity (Holtermann et al. 2018). Consequently, adjusting the model for occupation-attributed physical activity may give better meaning and improve the statistical significance of the physical activity variables.

Covariates have a determinant role in the model’s prediction accuracy and results significance. For model adjustment, we chose the covariates age categories, BMI categories, and CCI as a marker of
morbidity. We turned age into a categorical variable because of the sparse distribution of the subjects’ ages. Nevertheless, the distribution of the study subjects by age categories remained largely uneven. The age category 52-57 had the larger number of subjects (1152) and, when stratifying by age, it was sometimes the only stratum showing statistically significant results. Age is an undeniable confounding factor and its inclusion as a covariate is usually a non-questionable practice in medical research. The inclusion of morbidity measurements in the model as covariates on the other hand was function of various considerations many of which were supported by the literature review.

CCI, as a score meant to reflect the combined effects of multiple morbidities, defaulted in accuracy and objectivity when measured through self-reported questionnaires at baseline. Although that baseline CCI was found to correlate with an appropriately calculated CCI based on ICD-9 and ICD-10 codes from hospital registries data further in time, proper clinical data on the study subjects at baseline would have more objectively reflected their initial morbidity status. Nevertheless, CCI at baseline was found to significantly increase the risk of mortality (HR = 1.14, p-value < 0.001). Initially having comorbidities seems to substantially affect future outcomes. However, due to model fitness considerations, CCI was dropped off the smoking-stratified model afterwards.

Other factors have tremendous effects on mortality curves and might be essential to adjust for in order to get better model fitness and accuracy; not only at baseline, but also throughout the course of the follow-up. Under the circumstances of our study, we give here the example of retirement as an important modifier of physical activity as a behavior, and mental status as a comorbidity (Dave et al. 2006, Soldo et al. 2006). As a major event at mid-life, retirement affects subjects differently depending on the nature of their jobs, their socio-economic conditions, and other factors. The loss of a close relative is an event of similar eminence starting to rise in frequency at mid-life. A better consideration of the socio-economic and environmental circumstances of the individuals along the duration of follow-up might allow a deeper understanding of how behavior affects mortality and how it changes in patterns through time.

For instance, establishing clearer criteria of risk factors and covariates choice could have potentially improved the study’s methodology as it was done in Manuel et al.’s work on the burden of unhealth behaviors in Canada (2016). The team of researchers consulted policy actors to identify
different population subgroups as to maximize the process of consideration of exposures to behavioral risk. They, then, defined a set of criteria for the inclusion of risk factors as to determine what risk is clinically important for future planning and policy.

6.1.1 The BMI controversy

BMI is also an important covariate and determinant of obesity as a key-type of comorbidity. BMI was treated as a categorical variable using WHO classification of overweight and obesity. However, we had to join underweight with normal weight because of the small sample size of the former category (a few cases only). And we decided to divide overweight into two strata as to allow better model fitness.

While obesity was found to significantly increase the risk of mortality in comparison to normal weight (HR = 1.40, p-value <0.001), to our great surprise, the first stratum of overweight (BMI from 25 to 27.5) was found to be associated with better survival than the normal weight category in smokers (HR = 0.74, p-value 0.015) (Figure 15). However, in nonsmokers, slight overweight appeared statistically significantly associated with increased mortality (HR = 1.33, p-value = 0.017).

The slight overweight paradox, probably more commonly called the BMI controversy, could be explained by the fact that eating slightly more than needed guarantees enough storage of necessary nutrients. Another explanation could be that the excess weight is not always fat. Combining the level of physical activity with BMI might better reflect the nature of extra weight as athletes tend to have more muscular mass than people with lower levels of physical activity. Using other measures of body morphology and markers of abdominal obesity might improve the accuracy of the measurement of the effect of metabolic stress exercised by fat on the cardiovascular system in a better way (Hainer & Aldhoon-Hainerová 2013). It is also worth mentioning that many of the overweight might be already on the trajectory of becoming obese. However, what seems to be a more tangible explanation is that tobacco smoking has a quantity-dependent anorexic effect, and the more a subject smokes, the leaner they might be. Smokers with normal weight might be smoking a higher quantity of tobacco than overweight smokers, and thus, the quantity of tobacco smoked could be a confounder affecting both weight and mortality. Consequently, stratification by the quantity of tobacco smoked might clarify the situation.
6.1.2 The most important predictors of life expectancy

We have initially set the secondary objective of our study to determine which health behaviors are the most important predictors of life expectancy among the studied factors. However, the prediction models that we ended-up having for our study do not allow equitable comparison between the effect sizes of different variables. For instance, while smoking comes as a categorical variable with two distinct levels, BSDS comes as a continuous variable with no clear cut between a healthy and an unhealthy diet.

Continuous variables are also not directly comparable since units of input are different in nature. Proper comparison might be difficult even with what seem to be equivalent categorical variables. For example, if we consider transforming alcohol consumption into similar categorization as smoking (i.e. abstainer, alcohol consumer), it will not be fair to directly compare the effect size of smoking to the effect size of alcohol consumption. The harmful effects of alcohol consumption might not start to show significance until far after level zero while what is considered low level of tobacco smoking might already start showing significantly harmful health effects.

Moreover, the distribution of the smokers over different amounts of tobacco smoking might not be comparable to the distribution of the alcohol consumers over different levels of alcohol consumption. If most of alcohol consumers in the study are very light drinkers while most of the smoking study subjects are heavy smokers, the effect of alcohol drinking as a behavior might be underestimated and the results of the study will not be generalizable.

However, as an attempt to evaluate the contrasts between our selected health behaviors under a common framework, we have compared the predicted survival of a generated subject with ideal health attributes vs the predicted survival of generated subjects with similar ideal health attributes except for one unhealthy behavior – one behavior at a time. Unhealthy behaviors were set to the value of the third quartile of the respective factor in our population. The results (Figure 19) suggest that smoking is the most important health behavior in term of survival with nearly 7 attributed years lost followed by diet with over 4 years lost and then alcohol drinking with only 1 year lost.

The purpose of ranking health behaviors in term of importance might help determine health policy priorities for decision makers and might serve as a guide for health promotion resources allocation.
and target assignment. These interventions would ultimately lead to a change of behavior. Nevertheless, the ease of change of behavior might vary significantly from health behavior to another and from individual to individual as a function to many factors. Therefore, it is worthy to evaluate the likelihood to change behavior as another indicator of health behaviors importance. For instance, if it is more likely for an individual to sustainably change from a BSDS of 5 to a BSDS of 25 than to move from a status of smoker to a status of nonsmoker from example, it might be wiser to give more importance to changing the individual’s diet than smoking status. The likelihood of relapse might also need to be taken into consideration. Similarly, the individual’s circumstances, society, and environment have also a determining role in the ease of change. In a society where alcohol drinking is very common, it might be more difficult for an individual to abstain from drinking alcohol than to abstain from smoking tobacco.

7 CONCLUSION AND FUTURE RESEARCH

A simple model with which to predict life expectancy was presented. Survival can be predicted through a few easily obtainable health behavior measurements. The study provides evidence from the KIHD cohort on the tremendous effects of health behaviors on life-expectancy. As health information in general tend to be complicated for the general public and difficult to project to real life circumstances, one of the aims of this study is to use life-expectancy as a simple form to present the cumulative risk of the main lifestyle-related risk factors.

Smoking was found to be responsible of the loss of about 8 years of life. Up to 20 years of life are to be gained by adopting an optimal healthy lifestyle from midlife on. As the literature suggests, improving life expectancy would not only lead to a longer life, but also to a better quality of life (Fries 1980).

Life expectancy has often been used as a health indicator to estimate the overall status of health in a given population and guide policies to tackle inequity in sub-populations with different life expectancies. Results from our study can be used as a mean of risk communication permitting to properly present to individuals at risk the negative health outcomes caused by their behavior with the aim to induce them to change their behavior and reduce this risk (Gamhewage 2014).
This work also prepares the grounds for an online personalized risk assessment that can be used as a communication tool for health promotion. Such tool can also be refined to serve as a virtual indicator of life expectancy in different regions.

In addition to the correction of the methodologic weaknesses previously discussed, the study might benefit from new computer-based methods of prediction such as machine learning – as illustrated on the project framework. Moreover, further research could be done to assess how the social determinants of health as well as social-related factors such as social support, social interactions, and social inclusion, as the literature suggests it (Holt-Lunstad 2018, Holt-Lunstad et al. 2010, 2015, Teoh & Hilmert 2018, Valtorta et al. 2016), could affect our estimate of life expectancy.

8 REFERENCES


al-Baghdadi AM. التكملة في الحساب. 1020.


Ashrafian H. Familial stroke 2700 years ago. Stroke 2010;41:e187; author reply e188.


Eknoyan G. A history of obesity, or how what was good became ugly and then bad. Adv Chronic Kidney Dis 2006;13:421–427.


Emberson JR, Bennett DA. Effect of Alcohol on Risk of Coronary Heart Disease and Stroke: Causality, Bias, or a Bit of Both? Vasc Health Risk Manag 2006;2:239–249.


Grøntved A, Hu FB. Television Viewing and Risk of Type 2 Diabetes, Cardiovascular Disease, and All-Cause Mortality A Meta-analysis. JAMA 2011;305:2448–2455.


Porta M. Confounding by indication and past clinical trials. Epidemiology 1997;8:219–220.


identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 2012;44:991–1005.


