Socioeconomic status and health: the role of lifestyle choices

Sosioøkonomisk status og helse: betydningen av livsstilsvalg

Arnstein Øvrum

Norwegian University of Life Sciences • Universitetet for miljø- og biovitenskap
Department of Economics and Resource Management
Philosophiae Doctor (PhD) Thesis 2012:04

Arnstein Øvrum
Philosophiae Doctor (PhD) Thesis 2012:04

ISBN 978-82-575-1042-8
ISSN 1503-1667

NILF
Norwegian Agricultural Economics Research Institute

Norwegian Agricultural Economics Research Institute
P.O. Box 8024 Dep
NO–0030 OSLO, Norway
Phone +47 22 36 72 00
www.nilf.no, e-mail: postmottak@nilf.no

Norwegian University of Life Sciences
NO–1432 Ås, Norway
Phone +47 64 96 50 00
www.umb.no, e-mail: postmottak@umb.no

Socioeconomic status and health: the role of lifestyle choices

Sosioøkonomisk status og helse: betydningen av livsstilsvalg

Philosophiae doctor (PhD) Thesis

Arnstein Øvrum

Department of Economics and Resource Management
Norwegian University of Life Sciences

Ås 2011

Thesis number 2012:4
ISSN 1503-1667
ISBN 978-82-575-1042-8

Abstract

This thesis focuses on the relationship between socioeconomic status, health and related lifestyle choices. It uses repeated cross-section and stated preference data from Norway and consists of four papers.

The first paper compares sources of inequality in health, represented by self-assessed health and obesity, with sources of inequality in lifestyle choices central to the production of health, represented by physical activity, cigarette smoking and two indicators of healthy dietary behavior; the consumption of fish and the consumption of fruits and vegetables. The results demonstrate that patterns of inequality in health are not necessarily representative of patterns of inequality in important, underlying production factors of health, and that education and income are not always the most important sources of inequality in lifestyles.

The second paper examines how education and income differences in physical activity, the consumption of fruits and vegetables, cigarette smoking and self-assessed health evolve over the adult life course. Although mixed, the results provide some evidence of increased health consciousness and associated lifestyle improvements in age among lower socioeconomic status groups. Such improvements may potentially contribute to reducing cumulative advantage effects in health by socioeconomic status at older ages.

The third paper estimates the demand for physical activity and fruits and vegetables using latent class models, focusing on subpopulation heterogeneity in the effects of education and income. The results suggest that among the majority of the population that should be more physically active and eat more fruits and vegetables, the role of education and income may be even more important than previously assumed.

The fourth paper uses stated preference data on semi-hard cheese to examine how diet choices are affected by exposure to health information, and more specifically it examines to what extent such health information effects vary by education, income, age and gender. The

results suggest a promising role for health information policies in reducing educational differences in diet-health knowledge and thus dietary behavior. Targeting low income groups, young people and particularly males through health information policies seems more difficult.

Although the nature of our data do not allow for making causal inference, the results of this thesis are generally suggestive of there being a closer triangular relationship between education, lifestyles and health than between income, lifestyles and health. Thus, at least for policies aimed at improving population health through improved lifestyle habits, it seems more important to target low education groups than low income groups. Related to this, the results of the thesis demonstrate that one should be careful in treating socioeconomic status as a unified concept. Finally, although this thesis focuses mainly on the role of socioeconomic status, its results suggest that it in order to effectively improve overall population health, policy instruments for improved lifestyle habits should also consider the role of other and in some cases perhaps more important socio-demographic factors, including in particular age and gender.

Sammendrag

Denne avhandlingen fokuserer på sammenhengen mellom sosioøkonomisk status, helse og tilhørende livsstilsvalg. Den benytter repeterte tverrsnittsdata og eksperimentelle data fra Norge og består av fire artikler.

Den første artikkelen sammenligner kilder til ulikhet i helse, representert ved egenvurdert helse og fedme, med kilder til ulikhet i helserelaterte livsstilsvalg, representert ved fysisk aktivitet, røyking og to indikatorer på et sunt kosthold; etterspørselen etter fisk og etterspørselen etter frukt og grønnsaker. Resultatene tyder på at ulikhetsmønstre i helse ikke nødvendigvis er representative for ulikhetsmønstre i viktige, underliggende produksjonsfaktorer for helse, og at utdanning og inntekt ikke alltid er de viktigste kildene til ulikhet i livsstilsvalg.

Den andre artikkelen undersøker hvordan utdannings- og inntektsforskjeller i fysisk aktivitet, etterspørselen etter frukt og grønnsaker, røyking og egenvurdert helse utvikler seg over det voksne livsløpet. Resultatene er ikke entydige, men peker til en viss grad i retning av at personer i lavere sosioøkonomiske grupper blir mer helsebevisste når de blir eldre, med tilhørende forbedringer i livsstilsvaner. Slike forbedringer kan potensielt bidra til å redusere akkumuleringen av sosioøkonomiske ulikheter i helse over livsløpet.

Den tredje artikkelen estimerer etterspørselen etter fysisk aktivitet og frukt og grønnsaker ved bruk av latente klassemodeller, og fokuserer på subpopulasjonsheterogenitet i effektene av utdanning og inntekt. Resultatene tyder på at blant majoriteten av befolkningen som burde være mer fysisk aktive og spise mer frukt og grønnsaker, så kan betydningen av utdanning og inntekt være enda viktigere enn tidligere antatt.

Den fjerde artikkelen benytter data fra et valgeksperiment på gulost for å undersøke hvordan det å bli eksponert for helseinformasjon påvirker kostholdsvalg, og mer konkret fokuserer den på i hvilken grad slike helseinformasjonseffekter varierer på tvers av utdanning,

inntekt, alder og kjønn. Resultatene tyder på at helseinformasjon kan bidra til å redusere utdanningsforskjeller i kostholdskunnskap og dermed i kostholdsvaner. Å nå ut til lavinntektsgrupper, yngre mennesker og spesielt menn gjennom helseinformasjonstiltak synes vanskeligere.

Selv om begrensninger i vårt datamateriale ikke tillater etablering av kausale sammenhenger, så peker resultatene i avhandlingen generelt i retning av at det er en tettere triangulær sammenheng mellom utdanning, livsstilsvalg og helse enn mellom inntekt, livsstilsvalg og helse. Med hensyn til politiske tiltak som har som målsetning å forbedre folkehelsen gjennom forbedrete livsstilsvaner synes det derfor viktigere å nå ut til grupper med lav utdanning enn grupper med lav inntekt. De delvis ulike resultatene for utdanning og inntekt i avhandlingen tyder videre på at man bør være forsiktig med å anse sosioøkonomisk status som et felles, enhetlig begrep. Og til slutt, selv om denne avhandlingen primært fokuserer på betydningen av sosioøkonomisk status, så tyder dens resultater på at for å forbedre folkehelsen på en mest mulig effektiv måte, så bør man ved utforming av politiske tiltak for forbedrete livsstilsvaner også vurdere betydningen av andre og i noen tilfeller kanskje viktigere sosio-demografiske faktorer, herunder spesielt alder og kjønn.

Aknowledgements

This thesis work has been carried out at the Norwegian Agricultural Economics Research Institute (NILF). I would like to thank NILF and in particular Sjur Spildo Prestegard for generous support and patience throughout the thesis process. I thank all my colleagues at NILF for making it a good place to work, and I especially thank Geir Gustavsen, who is my closest collaborator at NILF and co-author of one of the papers in the thesis.

I am especially grateful to my main supervisor, Kyrre Rickertsen, for his excellent guidance, for always giving constructive feedback, and for helping me stay motivated during difficult periods of the thesis process. He is also co-author of three of the papers. I am also very thankful to my co-supervisor, Frode Alfnes, for sharing his insights on experimental methods and for his valuable feedback and suggestions. He is also co-author of one of the papers. Thanks also to the Department of Economics and Resource Management at the Norwegian University of Life Sciences for administrative support, and to the Development and Resource Economics group for their inspiring Master's programme, which triggered my interest in pursuing PhD studies and academic work.

I would also like to thank the Department of Economics at Iowa State University, and in particular Wallace E. Huffman, for hosting me as a visiting student during the academic year 2007–2008. Their excellent graduate courses laid a good foundation for my later thesis work, and I am grateful for having met so many friendly and talented fellow PhD students from all over the world.

I also learned a lot from my stay at Nofima Mat during fall 2008. I would especially like to thank Valérie Lengard Almli, Margrethe Hersleth and Tormod Næs. Valérie is also co-author of one of the papers.

I gratefully acknowledge the Research Council of Norway for funding my PhD scholarship and our research (grant numbers 182289, 184809 and 178200). I would also like to thank Synovate Norway for collecting the two data sets that are being used in this thesis, and I am also grateful to all the people who took their time to answer these surveys.

Finally I would like to thank my wife, Mette, for your love, patience and support. I look forward to spending less time writing papers and spending more time with you.

Oslo, December 2011

Arnstein Øvrum

Table of contents

Introduction and summary

This thesis focuses on the relationship between socioeconomic status, health and related lifestyle choices. It uses repeated cross-section and stated preference data from Norway and consists of four papers, which may be read independently. The first paper compares sources of inequality in health, represented by self-assessed health and obesity, with sources of inequality in lifestyle choices central to the production of health, represented by physical activity, cigarette smoking and two indicators of healthy dietary behavior; the consumption of fish and the consumption of fruits and vegetables. The second paper investigates the role of lifestyle choices, represented by physical activity, the consumption of fruits and vegetables and cigarette smoking, in explaining how education and income differences in self-assessed health evolve over the adult life course. The third paper estimates the demand for physical activity and fruits and vegetables using latent class models, focusing on subpopulation heterogeneity in the effects of education and income. The fourth paper uses stated preference data on semi-hard cheese to examine how diet choices are affected by exposure to health information, and more specifically it examines to what extent such health information effects vary by education, income, age and gender.

I will next briefly review the large literature on the demand for health and related lifestyle choices, focusing on the role of socioeconomic status. Following that I summarize and discuss the results, implications and limitations of the four papers of this thesis.

Background and motivation for the thesis

Most indicators of health and related lifestyle choices are unequally distributed within populations.[1] Sources of inequality in lifestyles and health include for example genetic

[1] Throughout this thesis, lifestyle choices (or just lifestyles) refer to everyday behaviors that may affect health (for example cigarette smoking). In the health inequality literature, health affecting lifestyles are also frequently referred to as health behaviors. This term will also be used some places in this thesis. Lifestyle choices may be regarded as a subset of a larger set of behaviors that may affect health.

disposition, age, gender, marital status, and more external factors such as physical work conditions and local water quality. However, studies on inequality in lifestyles and health have predominantly focused on the role of financial and human resources (van Doorslaer and van Ourti, 2011). These resources are most often represented by income and education, respectively, and by similar measures of our parents during childhood. These and similar sources of inequality, such as wealth, occupation and subjective social status (Cutler *et al.* 2006), are often referred to collectively as socioeconomic inequalities, or socioeconomic gradients, in lifestyles and health.[2]

The health inequality literature tends to focus on the role of socioeconomic status for several reasons. First, the positive correlation between socioeconomic status and good health is very strong and consistent; it holds almost irrespective of how socioeconomic status and health is defined, and it is found in all types of countries, including strong welfare states.[3] For example, in the complete cohort of Norwegians aged 45–64 years in 1990, mortality during the period 1990–1999 was almost two times higher among males with only lower secondary education (9 years) or less than among males with at least some university or college education (Strand *et al.*, 2010).[4] Second, although debated, many people including key policy makers argue that health differences by socioeconomic status are unfair and should be combated (Norwegian Ministry of Health and Care Services, 2006; CSDH, 2008).[5] And third,

---

[2] The term socioeconomic gradient in lifestyles and health is widely used and refers to the observation that, frequently, there is a close-to-linear relationship between on the one hand indicators of socioeconomic status and on the other hand indicators of good health and healthy lifestyles.

[3] This thesis will not address the literature that compares health inequalities across different rich countries, where central topics include the relationship between income inequality and health inequality and the role of educational systems and health care financing. Although highly important, this thesis will also not cover the issue of health differences across poor and rich countries. For a summary of these issues, see for example CSDH (2008).

[4] The mortality rates in these two education groups were 1,425 and 780 per 100,000 person years, respectively. The corresponding mortality rates among females were 726 and 426 per 100,000 person years (Strand *et al.*, 2010).

[5] Some studies use the term inequity in health '… for those inequalities in health that are deemed to be unfair or stemming from some form of injustice' (Kawachi *et al.*, 2002). For example, health inequalities directly attributable to maternal health behaviors during pregnancy (Barker, 1997) and country of birth (CSDH, 2008) are clearly unfair. On the other hand, health inequalities attributable to factors involving at least some element of personal choice or preference – including education and income – are more difficult to label as either totally fair

the relationship between socioeconomic status, health and related lifestyles is extremely complex and multifaceted, and much research is therefore needed to better understand these issues.

In economics, the literature on socioeconomic inequalities in health has mainly been driven by exploratory, empirical studies, although both educational attainment and income (or wages and wealth) are important elements of, for example, the human capital model of the demand for health (Grossman 1972, 2000). In this model, which is too comprehensive to be presented in detail here, health is viewed as both a consumption good (not being sick increases utility) and an investment good (not being sick increases available time for productive activities such as earning incomes). The individual inherits an initial stock of health that deteriorates in age. However, this health stock can to some extent be maintained or increased through relevant health investments such as purchasing medical care and choosing healthy lifestyles. Higher levels of education are assumed to increase the efficiency of health production through for example improving one's ability to process health information and take advantage of new health technologies. On the other hand, a higher wage rate increases the opportunity cost of time, which increases the incentive to stay in good health but, at the same time, makes time-consuming health investments such as physical activity relatively more expensive (because of higher foregone earnings). Thus, although overall the human capital model of the demand for health suggests that there should be a positive relationship between higher socioeconomic status and good health, it does not necessarily suggest the same for all types of good health investments.

The relationship between socioeconomic status, health and related lifestyle choices is probably too complex and multifaceted to be fully captured in one theoretical model such as the human capital model of the demand for health (Cutler *et al.*, 2011). Much of this

---

or totally unfair. This thesis will generally not draw clear distinctions between what are fair and unfair inequalities in health or try to define what these are. For thorough treatments on health inequality and fairness, see for example Fleurbaey and Schokkaert (2009) and Olsen (2011).

complexity revolves around the issue of causality versus correlation, including the direction of causal effects. That is, does higher socioeconomic status cause better health and healthier lifestyles, and if so why? Does better health cause higher socioeconomic status, and if so why? Or are socioeconomic status, lifestyles and health strongly correlated mainly because they are all strongly influenced by underlying, 'third' factors such as inborn cognitive and noncognitive skills? A complete review of the large and mainly empirical literature that addresses these important questions, where many issues still remain unresolved, is not given here due to space considerations. However, an overview of some of the leading and most frequently studied hypotheses is provided in Table 1.

Table 1
Socioeconomic status, lifestyle choices and health. Some possible mechanisms.

Causality that may run from socioeconomic status to health, possibly through lifestyles	Causality that may run from health to socioeconomic status, possibly through lifestyles	Possible 'third' factor explanations
▪ Education may increase the efficiency of health production through, for example, improving one's cognitive skills, including the ability to process and adapt to health information and new health technologies (Grossman, 1972, 2000; Cutler and Lleras-Muney, 2007) ▪ Education may affect noncognitive skills such as time preferences, risk averseness and self-control, which in turn may affect current lifestyle choices in the interest of future health and longevity (Fuchs, 1982; Cutler and Lleras-Muney, 2007) ▪ Higher wages (income) increases the opportunity cost of time and may thus lead to less engagement in time-consuming health investments such as physical activity (Grossman, 1972, 2000) ▪ Higher incomes make healthy lifestyle habits such as eating nutritious foods easier affordable (Blaylock *et al.* 1999) ▪ Low absolute and relative socioeconomic status may cause psychosocial stress, which may impact health both directly and through unhealthy lifestyle habits (Cutler *et al.*, 2006) ▪ Parental socioeconomic status may, through parental behavior, affect child health (including birth weight) and her/his health later in life (Case *et al.*, 2002; Currie, 2009)	▪ Poor health during adulthood that is not caused by low socioeconomic status (e.g., random health shocks due to genetic disposition) may lead to premature exits from the labor force, which in turn lowers income due to a shift from earning wages to being on social security (Case and Deaton, 2005) ▪ Poor health during childhood (including low birth weight) and adolescence that is not caused by parental socioeconomic status (e.g., random health shocks) may directly affect educational attainment and earnings later in life (Case *et al.*, 2005; Black *et al.*, 2007; Currie, 2009)	▪ Inborn cognitive and noncognitive skills may affect educational attainment and later earnings, and be correlated with adult lifestyle habits (and thus adult health) (Heckman, 2006, 2007; Conti *et al.*, 2010; Cutler and Lleras-Muney, 2010) ▪ Differences in lifestyle habits (and thus health) across different education and income groups may reflect the communication of group membership and social identity (Akerlof and Kranton, 2000; Etilé, 2007) ▪ Low education/income jobs often involve manual, physically strenuous work, which may impact health negatively in the long run (Case and Deaton, 2005)

As suggested by some of the listed hypotheses in Table 1, economists are increasingly looking to other disciplines such as psychology and sociology to gain a better and more complete understanding of why indicators of socioeconomic status and health are so closely related. However, for factors such as for example self-control and time preferences, it is generally difficult to determine to what extent these influence educational attainment in the first place, and to what extent they are influenced by the education process itself. Recent studies by for example Heckman (2006, 2007) and Conti *et al.* (2010) underscore the dynamic nature of cognitive and noncognitive skill formation, where childhood and adolescent years are particularly important.

These studies are part of an increasing literature that focuses on the importance of childhood health and circumstances in affecting similar outcomes in adulthood, including the issue of intergenerational transmission of socioeconomic inequalities in health. Lower birth weight and poor health during childhood is associated with lower educational attainment, lower earnings and poorer health in adult life, even among twins and siblings (Barker, 1997; Case *et al.*, 2005; Black *et al.*, 2007; Currie, 2009). Parental education and income is significantly associated with the child's health and with her or his socioeconomic status, lifestyle habits and health later in life (Case *et al.*, 2002; Currie, 2009; Rosa Dias, 2010). Although arguably important, the role of childhood health and circumstances in affecting similar outcomes in adulthood adds to the complexity of understanding and disentangling the many sources that produce socioeconomic inequalities in health.

Several studies have utilized data from natural experiments in the form of, for example, school reforms to examine how health is affected by exogenous variation in length of education. A majority of these studies seem to confirm that there are at least some casual effects running from higher education to better health and healthier lifestyles (Lleras-

Muney, 2005; van Kippersluis *et al.*, 2011; Cutler *et al.*, 2011). Furthermore, higher maternal education causes better child health (Currie and Moretti, 2003).

Another group of studies have examined how education and income differences in health evolve over the adult life course. These studies have shed light on some of the fundamental differences between education and income. For example, education is more or less predetermined in such a setting while income may be affected by many factors throughout the adult life course, including health shocks and the gradual deterioration of health (Smith, 2004). The correlation between income and health is often found to be particularly strong during some of the last years before expected retirement, and this seems to largely reflect the effect of poor health on premature exit from the labor force. Thus, poor health affects incomes negatively due to a shift from earning wages to being reliant on social security payments (Case and Deaton, 2005; van Kippersluis *et al.*, 2010). Also observations from other types of studies suggest that the causal effects of higher income on health during adulthood may be relatively small, at least above some minimum income level (Cutler *et al.*, 2011). These studies include studies that examine the effects of economic recessions on lifestyle habits and health (Ruhm, 2000, 2005), and studies that examine the effects of lagged income on the onset of new health conditions (Smith, 2007). Thus, as illustrated by this brief literature review and noted by Cutler *et al.* (2011), it might be misguided to treat socioeconomic status as a unified concept. Socioeconomic status consists of many dimensions, including education, income, occupation, self-perceived social status and so on, and these dimensions relate to health in diverse ways.

Numerous studies have reported significant associations between socioeconomic status and health affecting lifestyles such as physical activity, cigarette smoking and dietary behavior. However, these and similar lifestyle choices have received relatively little explicit attention in the health inequality literature. Where considered, lifestyles have usually played a secondary role in that they are regarded as part of a larger set of factors that produce total inequalities and socioeconomic inequalities in health (Balia and Jones, 2008; Costa-Font and Gil, 2008). Few studies have gone one step further and examined what are the key sources of inequality in health affecting lifestyles themselves, including to what extent these sources are the same as in health. For example, is income equally important in explaining inequality in cigarette smoking as in self-assessed health? Do the education and income gradients in lifestyles remain constant throughout the adult life course? If so, this would suggest that the corresponding gradients in health should be gradually increasing in age due to the long-term, cumulative nature of health production (Kim and Durden, 2007). On the other hand, if people in lower socioeconomic status groups grow more health conscious as they age and improve their lifestyle habits accordingly, this may contribute to reducing such cumulative advantage effects in health by socioeconomic status at older ages. Examples of studies that partly address these questions using data from the US and the UK include Cutler and Lleras-Muney (2007, 2010) and Cutler *et al.* (2011). The two first papers of this thesis address similar questions, but using Norwegian data and somewhat different methodological approaches.

Although most empirical studies confirm *a priori* expectations about positive effects of higher education and income on healthy lifestyles, the marginal effects derived from conventional mean-effects type econometric models are sometimes small or imprecisely estimated (Variyam *et al.*, 2002; Contoyannis and Jones, 2004). One of several possible explanations for such small marginal effects is so-called subpopulation heterogeneity; in certain segments or groups of the population, education and income may not be so closely associated with healthy lifestyles. Such preference heterogeneity may be accommodated using latent class models (Cameron and Trivedi, 1998). In the health economics literature, latent class models have mainly been applied on data for health care utilization (Deb and Trivedi, 1997, 2002; Bago d'Uva, 2005; Hole, 2008). The third paper of this thesis explores the use of latent class models in the context of lifestyle choices, focusing on subpopulation heterogeneity in the effects of education and income.

In many countries including Norway, reducing health differences by socioeconomic status is a key health policy goal (Norwegian Ministry of Health and Care Services, 2006; CSDH, 2008). Numerous policy instruments – direct and indirect, preventive and treatment-oriented – may impact health and the distribution of health across different socioeconomic groups. Policies that indirectly may prevent poor health and reduce socioeconomic inequalities in health include full kindergarten coverage, close-to-free educational and health care systems, and income redistribution through progressive tax systems and generous social security schemes (Norwegian Ministry of Health and Care Services, 2006). Direct policies for improving people's lifestyle habits and thus prevent poor health include price policies, restrictions and regulations, and dissemination of health information. For example, diet-related health information may help improve knowledge, raise awareness, reduce confusion and thereby make healthier food options more attractive and visible. The distributional effects of such health information policies across socio-demographic groups are difficult to measure and thus not well-known. For example, how will different education groups respond to a public information campaign on the importance of following a healthy diet? Due to different *a priori* levels of diet-health knowledge, it seems reasonable to expect that the marginal effects of health information on preferences for healthy foods should be larger in lower than higher education groups. On the other hand, low and high education groups may be systematically different in their ability to process and adapt to health information (Grossman, 2000), as well as in their general interest for health information. Thus, the effects of health information may also be positively associated with years of schooling. The fourth paper of this thesis explores the distributional effects of health information using data from a stated preference experiment on semi-hard cheese.

The thesis

The general objective of this thesis is to add to the existing literature on socioeconomic inequalities in health. It does so by mainly focusing on inequalities in the following important, underlying production factors of health; physical activity, cigarette smoking and three indicators of healthy dietary behavior; the consumption of fruits and vegetables, the consumption of fish, and preferences for low-saturated-fat and low-fat cheese. These lifestyle indicators are closely associated with the risk of major health outcomes, including type II diabetes, cardiovascular disease and certain types of cancer (World Health Organization, 2003). As health indicators this thesis focuses mainly on self-assessed health, which has been shown to be strongly correlated with several objective health measures (Idler and Benyamini, 1997), and to a lesser extent obesity, which is an intermediate risk factor for chronic diseases and itself a direct cause of reduced physical and mental health.

In line with most of the literature on socioeconomic inequalities in health, the four papers of the thesis are exploratory and empirically oriented. The three first papers use data from the Norwegian Monitor Survey, which is a nationally representative and repeated cross-section survey of adults aged 15–95 years. The survey has been conducted every second year since 1985, and the 3,000–4,000 respondents in each survey round answer an extensive list of questions on a wide range of topics. The fourth paper uses data from a stated preference experiment on semi-hard cheese. The experiment was part of an Internet survey that was conducted during spring 2009, and the responses of 408 participants are used in the paper.

The main objectives of the four papers are:

- To compare sources of total inequality and socioeconomic inequality across six important lifestyle and health indicators (Paper 1)
- To investigate the role of lifestyles in explaining how education and income differences in health evolve over the adult life course (Paper 2)
- To explore possible subpopulation heterogeneity in the demand for healthy lifestyles, focusing on the role of education and income (Paper 3)
- To examine how diet choices are affected by exposure to health information, and more specifically to examine how such health information effects vary by education, income, age and gender (Paper 4)

The four papers are described and summarized below. Following that I discuss some of the contributions, implications and limitations of the thesis.

Paper 1: Inequality in health vs. inequality in lifestyles
(co-authored with Kyrre Rickertsen)

Levels of inequality in health may be measured in several ways. In economics, levels of total inequality and socioeconomic inequality in health are usually measured using the Gini index and the concentration index, respectively. These inequality indices are particularly useful because of their decomposition properties (van Doorslaer and Jones, 2003). For example, one may calculate the percentage contribution of specific factors such as age, gender and parental education to total inequality and income-related inequality in self-assessed health. While many previous studies have used such decomposition techniques to investigate sources of inequality in health indicators such as self-assessed health, obesity and mortality, as well as the utilization of health care services, few, if any studies have undertaken similar assessments in key health affecting lifestyle choices.

This paper uses data from the Norwegian Monitor Survey 2005–2009 to directly compare sources of inequality in health, represented by self-assessed health and obesity, with sources of inequality in lifestyle choices central to the production of health, represented by physical activity, cigarette smoking and two indicators of healthy dietary behavior; the consumption of fish and the consumption of fruits and vegetables. As potential sources of inequality, we consider demographic factors, education, income, occupation, childhood circumstances, and proxies for time preferences, risk and self-control. Sources of inequality are compared by estimating a multivariate probit model for lifestyles and health, and by decomposing the explained part of the associated Gini indices and education- and income-related concentration indices.

The results of the decomposition analyses vary considerably across the three different inequality measures. Not surprisingly, education makes a substantial contribution to the explained part of the education-related concentration indices in lifestyles and health (mean: 67.9%), while income similarly makes a substantial contribution to the income-related concentration indices (mean: 49.6%). However, education and income are much less important in explaining total inequality in lifestyles and health, with mean Gini contributions of 18.4% and 10.0%, respectively. While education is found to be relatively important in explaining total inequality in all four lifestyle indicators (mean: 22.8%), income is relatively unimportant in fruits and vegetables consumption (3.9%), fish consumption (1.6%) and cigarette smoking (3.7%). In several cases, education and income are clearly outranked by other factors in terms of explaining total inequality, such as gender in fruits and vegetables consumption (47.8%), age in fish consumption (64.8%) and maternal education in obesity (20.9%). In summary, the results of this study suggest that patterns of inequality in health are not necessarily representative of patterns of inequality in important, underlying production factors of health.

Paper 2: Health inequalities over the adult life course: the role of lifestyle choices

(co-authored with Geir Wæhler Gustavsen and Kyrre Rickertsen)

Acknowledging the dynamic nature of health production, some studies in the health inequality literature have focused on how socioeconomic inequalities in health evolve over the adult life course (Case and Deaton, 2005; van Kippersluis *et al.*, 2010). This paper uses data from the Norwegian Monitor Survey 1997–2009 to explore the role of lifestyle choices in explaining these dynamics. Linear probability models are used to track income and education gradients in physical activity, the consumption of fruits and vegetables, cigarette smoking and self-assessed health over the age range 25–79 years. Sensitivity of the age-specific income and education gradients are assessed by the step-wise inclusion of additional control variables, including occupational status and a variety of socio-demographic characteristics.

While the education gradients in physical activity and the consumption of fruits and vegetables remain relatively stable throughout the adult life course, the education gradient in smoking is clearly decreasing in age. This life course pattern appears too pronounced to be explained fully by sample selection due to high rates of mortality among low-educated smokers, or by cohort effects due to, for example, the increasing stigmatization of cigarette smokers in recent decades. With the exception of the income gradient in physical activity among females, the income gradients in lifestyles are generally concave in age and decreasing slightly at older ages. At the same time, the role of lifestyles in moderating the relationship between income and self-assessed health appears modest. While the age-specific education gradients in self-assessed health are reduced by 27.8% on average when the three lifestyle indicators are added as control variables to the model, the corresponding income gradients are reduced by only 6.6%. This result partly reflects that while the income gradients in lifestyles are substantially reduced once we control for education, the reverse is not true.

The education and income gradients in subjective health consciousness are also examined. These are found to be gradually decreasing in age, and they actually turn from positive to negative at 64 years of age and remain negative thereafter. Thus, overall, while income and education differences in daily lifestyle choices should generally contribute to cumulative advantage effects in health by socioeconomic status over the adult life course, our results provide some evidence of increased health consciousness and associated lifestyle improvements in age among lower socioeconomic status groups. This could potentially contribute to reducing cumulative advantage effects in health by socioeconomic status at older ages. However, our results for education are too mixed and our results for income are too uncertain to firmly conclude on these matters.

Paper 3: Socioeconomic status and lifestyle choices: evidence from latent class analysis

(single-authored. This paper was published in Health Economics in 2011)

This paper uses data from the Norwegian Monitor Survey 1999–2009 to explore possible subpopulation heterogeneity in the demand for physical activity and the consumption of fruits and vegetables, focusing on the role of education and income. It does so by comparing results from conventional econometric count data models and their latent class model counterparts. In latent class models, the population is viewed as a probabilistic mixture of a finite set of subpopulations, or latent classes or groups of individuals (Cameron and Trivedi, 1998). In estimation, the log-likelihood function is specified as a weighted average of sub-distributions or component densities, of which each represent a different group or ‘type’ of individuals. Thus, the intercept and slope parameters – or the utility functions – are allowed to vary across groups, but is assumed fixed within each group. The weights, which are estimated along the component densities, reflect the average probabilities of belonging to the different groups.

For both physical activity and fruits and vegetables, the latent class models identify two subpopulations, or groups of people, with different sets of preferences. The minority groups, representing respectively 38.2% and 29.8% of the population, have high latent demands for physical activity and consumption of fruits and vegetables. In these groups, variability in demand is poorly explained by socioeconomic status. The two majority groups have low latent demands for healthy lifestyles, but in these groups, the marginal effects of higher education and income are generally much stronger than predicted by the conventional econometric count data models. Thus, for individuals in these important target groups for improved health, the socioeconomic gradient in important lifestyle choices may be steeper and thus more severe than previously assumed. Posterior analysis shows that individuals with higher socioeconomic status are more likely to belong to these healthier minority groups. Proxies for time preferences, risk, self-control and time constraints are also found to be important in characterizing these groups.

Paper 4: Health information and diet choices: results from a cheese experiment

(co-authored with Frode Alfnes, Valérie Lengard Almli and Kyrre Rickertsen)

Our daily decisions about eating healthy or unhealthy foods are influenced by a highly complex mix of factors. Nutrition policies may target at least two of these factors, health knowledge and awareness, through dissemination of diet-related health information. However, the distributional effects of such policies across socio-demographic groups are difficult to measure and thus not well-known. This paper utilizes properties of a controlled experiment to explore such distributional effects. In the stated preference experiment, which focuses on healthy attributes in semi-hard cheese, about half of the 408 participants were exposed to health information before performing either a choice or a ranking task. The effects of health information on marginal willingness to pay for low-saturated-fat, low-fat and organic cheese are analyzed using rank-ordered mixed logit models.

Overall, exposure to health information has a significant effect on marginal willingness to pay for low-saturated-fat and low-fat cheese. Furthermore, non-college, medium-high income, age 50–70 and female participants are more clearly affected by health information than college, low income, age 30–49 and male participants. In these former groups, the health information effects are always statistically at the 95% level, while in the latter groups they are generally not.

Subjective statements on diet-health knowledge and awareness are used to discuss these results. Based on ordered logit models, education is found to be a strong indicator of prior diet-health knowledge, but is simultaneously unrelated to the four statements on diet-health awareness. This finding corroborates well with the results in the cheese experiment; non-college participants learn relatively more from health information than college-educated participants, and they are therefore more clearly affected by it in the experiment. The statements on diet-health awareness are clearly associated with age and gender. Age 50–70 and female participants are more health conscious than age 30–49 and male participants, and this may explain why they are also more clearly affected by health information in the cheese experiment. Income is unrelated to both diet-health knowledge and awareness, but is clearly related to concerns about food prices. Thus, our finding that medium-high income participants are more clearly affected by health information than low income participants seems to mainly reflect the fact that the information effects are measured in terms of marginal willingness to pay, which is likely to depend in part on income and associated food budget constraints.

Contributions, implications and limitations of the thesis

The four papers of this thesis contribute to the literature on socioeconomic inequalities in health. The first paper demonstrates that patterns of inequality in health are not necessarily representative of patterns of inequality in important, underlying production factors of health. This suggests that the health inequality literature may benefit from paying more attention to patterns of inequality in factors of health production, including important lifestyle choices, in addition to health itself. The health inequality literature may also benefit from focusing more on sources of total inequalities in lifestyles and health, in addition to its current focus on sources of socioeconomic inequalities in health (Fleurbaey and Schokkaert, 2009; van Doorslaer and van Ourti, 2011). To efficiently improve overall population health and at the same time reduce the variance of health, one should search for key sources of population differences in single, important production factors of health, including different lifestyle choices, and in turn design tailored policies for each of these production factors. For example, health information on the importance of eating fish and eating fruits and vegetables could be targeted specifically towards young people and males, respectively. At the same time, education is found to be a relatively important source of inequality in all the considered lifestyle indicators of this study, and thus policies targeted specifically towards low education groups are also relevant.

The second paper illustrates that it is useful to consider the role of lifestyle choices in explaining how socioeconomic differences in health evolve over the adult life course. In both low and high socioeconomic status groups, our results generally point toward increased health consciousness and associated lifestyle improvements in age as a mechanism in slowing down the natural deterioration of physical health in age. However, as noted, our results for education are too mixed and our results for income are too uncertain to conclude that this process of 'compensating behaviour' at older ages is relatively stronger among lower than higher socioeconomic status groups. Thus, the role of dynamics in the relationship between socioeconomic status and lifestyles in either speeding up or slowing down cumulative advantage effects in health by socioeconomic status is not clear. Although income differences in lifestyles may play some role in explaining why there are income differences in health, including how these differences evolve over the adult life course, this seems less clear than in the case of education. Given that the education gradients in physical activity, consumption of fruits and vegetables and cigarette smoking are either stable or declining over the adult life course, policies for improved lifestyle habits should mainly target young people, and particularly young people with low levels of formal education. However, targeting these groups effectively through, for example, pricing and health information policies may be difficult. That said, our results suggest that in particular among low education groups, health consciousness is increasing in age. Thus, health information policies aimed towards making people more health consciousness at earlier stages of the adult life course may be efficient. Such health information could focus on the long-term, cumulative nature of health production and thus the importance of making healthy lifestyle choices also at younger ages.

The third paper contributes by introducing latent class models to the context of socioeconomic status and lifestyle choices in adults. Lifestyle choices are inherently complex; a wide variety of socio-demographic, psychological, psychosocial and institutional triggers and constraints affect whether we choose to live healthy. *A priori*, it is natural to expect that the more complex is the choice situation, the less homogeneous is the population. The results of the paper suggest a promising role for latent class models in accommodating preference heterogeneity associated with complex lifestyle choices. The empirical results of the paper suggest that among the majority of the population that should be more physically active and eat more fruits and vegetables, the role of education and income may be even more important than previously assumed. It seems that in conventional mean-effects type econometric models, the marginal effects of education and income are 'attenuated' as a result of socioeconomic status being unimportant among a healthy minority group of the population.

The fourth paper provides useful insights about the distributional effects of a specific policy instrument; the dissemination of diet-related health information. Our results suggest a promising role for health information polices in reducing educational differences in diet-health knowledge and thus dietary behavior. However, a challenge remains in how to effectively target low education groups in non-experimental settings. Also, according to our results, targeting low income groups, young people and particularly males through health information policies seems difficult. Experiences from smoking suggest that using ads or campaigns that contain personal stories or highly emotional elements such as films and images showing blocked blood vessels, tumors, heart attacks and so on that could result from years of cigarette smoking may be efficient in reaching young people and low socioeconomic status groups (Wakefield *et al.*, 2010). Such poor health outcomes could also result from years of excessive energy intakes and poor nutrition. Therefore, at least as a research exercise, it would be interesting to examine the distributional effects of similar, negatively loaded health information messages in the context of dietary behavior, and compare these with the distributional effects of more traditional, positively loaded health information initiatives such as ‘MyPlate’ in the US, the similar ‘Eatwell Plate’ in the UK, the ‘Keyhole’ labeling scheme in the Nordic countries and the ‘5 A Day’ campaign in various European countries.

The results and implications of this thesis must be viewed in light of its limitations. The three first papers use repeated cross-section data, and thus we are generally not able to make causal inference on the relationship between socioeconomic status, lifestyles and health, nor are we able to capture the dynamic nature of health production in single individuals. The fourth paper uses non-binding willingness to pay data from a stated preference experiment, and such data are often associated with hypothetical bias (Hensher, 2010). Thus, the results of this thesis must mainly be considered as representing new but tentative ideas and insights on the relationship between socioeconomic status, lifestyle choices and health. Both the data sources being used in this thesis are based on self-reported measures of socio-demographic characteristics, lifestyle habits and health. This may represent an additional source of error and bias, although for example self-assessed health has been shown to be strongly correlated with several objective health measures (Idler and Benyamini, 1997). Finally, for all the issues examined in this thesis, more similar research is needed before any firm conclusions can be drawn. Amongst others, this include similar studies using panel data, studies on other lifestyle indicators, studies on the same lifestyle indicators as in this thesis but with alternative variable definitions, studies from other countries, and field and lab experiments with consequential choices.

In conclusion, although the nature of our data do not allow for making causal inference, the results of this thesis are generally suggestive of there being a closer triangular relationship between education, lifestyles and health than between income, lifestyles and health. Thus, at least for policies aimed at improving population health through improved lifestyle habits, it seems more important to target low education groups than low income groups. Related to this, the results of the thesis demonstrate that one should be careful in treating socioeconomic status as a unified concept (Cutler *et al.*, 2011). Finally, although this thesis focuses mainly on the role of socioeconomic status, its results suggest that it in order to effectively improve overall population health, policy instruments for improved lifestyle habits should also consider the role of other and in some cases perhaps more important socio-demographic factors, including in particular age and gender.

References

Akerlof GA, Kranton RE. 2000. Economics and identity. *Quarterly Journal of Economics* **115**: 715–753.

Bago d'Uva T. 2005. Latent class models for use of primary care: evidence from a British panel. *Health Economics* **14**: 873–892.

Balia S, Jones AM. 2008. Mortality, lifestyle and socio-economic status. *Journal of Health Economics* **27**: 1–26.

Barker DJP. 1997. Maternal nutrition, fetal nutrition and diseases in later life. *Nutrition* **13**: 807–813.

Black SE, Devereux PJ, Salvanes KG. 2007. From the cradle to the labor market? The effect of birth weight on adult outcomes. *Quarterly Journal of Economics* **122**: 409–439.

Blaylock J, Smallwood D, Kassel K, Variyam J, Aldrich L. 1999. Economics, food choices and nutrition. *Food Policy* **24**: 269–286.

Cameron AC, Trivedi PK. 1998. *Regression analysis of count data.* Cambridge University Press: New York.

Case A, Deaton A. 2005. Broken down by work and sex: how our health declines. In *Analyses in the economics of aging.* Wise DA (Ed.). Chicago University Press: Chicago.

Case A, Fertig A, Paxson C. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics* **24**: 365–389.

Case A, Lubotsky D, Paxson C. 2002. Economic status and health in childhood: the origins of the gradient. *American Economic Review* **92**: 1308–1334.

Conti G, Heckman J, Urzula S. 2010. The education-health gradient. *American Economic Review* **100**: 234–238.

Contoyannis P, Jones AM. 2004. Socio-economic status, health and lifestyle. *Journal of Health Economics* **23**: 965–995.

Costa-Font J, Gil J. 2008. What lies behind socio-economic inequalities in obesity in Spain? A decomposition approach. *Food Policy* **33**: 61–73.

CSDH. 2008. Closing the gap in one generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health. World Health Organization: Geneva.

Currie J, Moretti E. 2003. Mother's education and the intergenerational transmission of human capital: evidence from college openings. *Quarterly Journal of Economics* **118**: 1495–1532.

Currie J. 2009. Healthy, wealthy and wise: socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature* **47**: 87–122.

Cutler DM, Deaton A, Lleras-Muney A. 2006. The determinants of mortality. *Journal of Economic Perspectives* **20**: 97–120.

Cutler DM, Lleras-Muney A, Vogl T. 2011. Socioeconomic status and health: dimensions and mechanisms. In *The Oxford handbook of health economics*. Glies S, Smith PC (Eds.). Oxford University Press: Oxford.

Cutler DM, Lleras-Muney A. 2007. Education and health: evaluating theories and evidence. In *The health effects of social and economic policy.* House JS, Schoeni RF, Kaplan GA, Pollack H (Eds.). Russell Sage Foundation: New York.

Cutler DM, Lleras-Muney A. 2010. Understanding differences in health behaviors by education. *Journal of Health Economics* **29**: 1–28.

Deb P, Trivedi PK. 1997. Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics* **12**: 313–336.

Deb P, Trivedi PK. 2002. The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics* **21**: 601–625.

van Doorslaer E, Jones AM. 2003. Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics* **22**: 61–87.

van Doorslaer E, van Ourti T. 2011. Measuring inequality and inequity in health and health care. In *The Oxford handbook of health economics.* Glied S, Smith PC (Eds.). Oxford University Press: Oxford.

Etilé F. 2007. Social norms, ideal body weight and food attitudes. *Health Economics* **16**: 945–966.

Fleurbaey M, Schokkaert E. 2009. Unfair inequalities in health and health care. *Journal of Health Economics* **28**: 73–90.

Fuchs V. 1982. Time preference and health: an exploratory study. In *Economic aspects of health*. Fuchs V (Ed.). University of Chicago Press: Chicago.

Grossman M. 1972. On the concept of health capital and the demand for health. *Journal of Political Economy* **80**: 223–255.

Grossman M. 2000. The human capital model of the demand for health. In *Handbook of health economics*. Culyer AJ, Newhouse JP (Eds.). Elsevier: Amsterdam.

Heckman JJ. 2006. Skill formation and the economics of investing in disadvantaged children. *Science* **312**: 1900–1902.

Heckman JJ. 2007. The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences* **104**: 13250–13255.

Hensher DA. 2010. Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B* **44**: 735–752.

Hole AR. 2008. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics* **27:** 1078–1094.

Idler EL, Benyamini Y. 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior* **38**: 21–37.

Kawachi I, Subramanian SV, Almeida-Filho N. 2002. A glossary of health inequalities. *Journal of Epidemiology and Community Health* **56**: 667–652.

Kim J, Durden K. 2007. Socioeconomic status and age trajectories of health. *Social Science & Medicine* **65**: 2489–2502.

van Kippersluis H, O'Donnell O, van Doorslaer E, van Ourti T. 2010. Socioeconomic differences in health over the life cycle in an egalitarian country. *Social Science & Medicine* **70**: 428–438.

van Kippersluis H, O'Donnell O, van Doorslaer E. 2011. Long-run returns to education: does schooling lead to an extended old age? *Journal of Human Resources* **46**: 695–721.

Lleras-Muney A. 2005. The relationship between education and adult mortality in the United States. *Review of Economic Studies* **72**: 189–221.

Norwegian Ministry of Health and Care Services. 2006. National strategy to reduce social inequalities in health. *Report No. 20 (2006-2007) to the Storting*: Oslo.

Olsen JA. 2011. Concepts of equity and fairness in health and health care. In *The Oxford handbook of health economics*. Glies S, Smith PC (Eds.). Oxford University Press: Oxford.

Rosa Dias P. 2010. Modeling opportunity in health under partial observability of circumstances. *Health Economics* **19**: 252–264.

Ruhm CJ. 2000. Are recessions good for your health? *Quarterly Journal of Economics* **115**: 617–650.

Ruhm CJ. 2005. Healthy living in hard times. *Journal of Health Economics* **24**: 341–363.

Smith JP. 2004. Unraveling the SES-health connection. *Population and Development Review,* Supplement: Aging, Health, and Public Policy **30**: 108–132.

Smith JP. 2007. The impact of social economic status on health over the life-course. *Journal of Human Resources* **42**: 739–764.

Strand BH, Grøholt E, Steingŕimsdóttir ÓA, Blakely T, Graff-Iversen S, Næss Ø. 2010. Educational inequalities in mortality over four decades in Norway: prospective study of middle aged men and women followed for cause specific mortality, 1960-2000. *British Medical Journal* **340:** c654 doi:10.1136/bmj.c654.

Variyam JN, Blaylock J, Smallwood D. 2002. Characterizing the distribution of macronutrient intake among US adults: a quantile regression approach. *American Journal of Agricultural Economics* **84**: 454–466.

Wakefield MA, Loken B, Hornik RC. 2010. Use of mass media campaigns to change health behaviour. *Lancet* **376**: 1261–1271.

World Health Organization. 2003. Diet, nutrition and the prevention of chronic diseases. *WHO Technical Report Series No. 916.* WHO: Geneva..

Paper 1

Inequality in health vs. inequality in lifestyles

Arnstein Øvrum[a,b*] and Kyrre Rickertsen[a,b]

[a]Department of Economics and Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway
[b]Norwegian Agricultural Economics Research Institute, P.O. Box 8024 Dep, N-0030 Oslo, Norway

Abstract

This paper uses Norwegian data to compare patterns of inequality in health, represented by self-assessed health and obesity, with patterns of inequality in lifestyle choices central to the production of health, represented by physical activity, smoking and diet quality. As potential sources of inequality, we consider demographic factors, education, income, occupation, psychological traits, and childhood circumstances. Patterns of inequality are compared by estimating a multivariate probit model for lifestyles and health, and by decomposing associated Gini and concentration indices. Heterogeneous patterns are revealed. Education is generally an important source of total inequality, while the role of income is mixed. In several cases, education and income are clearly outranked by other factors in terms of explaining inequality, such as gender in fruits and vegetables, age in fish consumption and maternal education in obesity. More studies that directly compare patterns of inequality in health production factors and health itself are needed.

JEL classification: D12; I12; I14; I18

Keywords: health; inequality; lifestyles; obesity; self-assessed health; socioeconomic status

---

[*] Corresponding author. E-mail: arnstein.ovrum@nilf.no. This paper was submitted to a peer-reviewed journal in June 2011.

1. Introduction

In recent years, considerable efforts have been made in order to improve our understanding of patterns of inequality in health, including attempts at identifying their sources. In particular, decomposition techniques for the Gini index and the concentration index (CI) have helped identify the relative contribution of individual characteristics and other factors to total inequality and socioeconomic inequality in health, respectively (van Doorslaer and Jones, 2003). Some studies have used these decomposition techniques to consider the role of health behaviors, or lifestyle choices, in explaining inequality in health, and in mediating the direct effect of socioeconomic status on health (Balia and Jones, 2008; Costa-Font and Gil, 2008; Vallejo-Torres and Morris, 2010).[1] However, few, if any studies have investigated sources of inequality in lifestyles themselves, rather than final health, using such decomposition techniques.

To reduce inequalities in health, and in particular socioeconomic inequalities in health, is stated as a key goal for health policy in many countries (CSDH, 2008). Policies that seek to address health inequalities may be most efficient when they are targeted towards the production factors of health, including lifestyles, and not final health itself. However, in order for such policies of 'preventive medicine' to be efficient, there is a need for more insights into patterns of inequality across several important health affecting lifestyles, including the extent to which these patterns are similar to those in final health. If patterns of inequality are homogeneous across lifestyles and health, it becomes relevant to use findings from studies on health – which is typically the focus of empirical work – as a basis for formulating policies for reduced inequality in lifestyles, due to the 'trickle down' properties from health to lifestyles.

[1] Using British panel data, Balia and Jones (2008) found that lifestyles explained about 25% of the variation in the Gini index for predicted mortality. Moreover, after allowing for endogeneity of lifestyles, the direct role of socioeconomic status in predicting mortality was reduced. Costa-Font and Gil (2008) found that physical activity, smoking, and food habits explained respectively 5.8%, 2.6% and 0.12% of the income-related CI in obesity in Spain. In a similar way, Vallejo-Torres and Morris (2010) found that smoking explained 2.3% of the income-related CI in health (as measured by EQ-5D) in England.

If on the other hand patterns of inequality vary significantly across different lifestyle and health variables, this no longer holds, which if true may have important implications for policy.

The objective of this paper is to compare patterns of inequality in health, represented by self-assessed health (SAH) and obesity, with patterns of inequality in lifestyle choices central to the production of health, represented by physical activity, smoking, and two indicators of diet quality. These are the frequency of consumption of fish, and the frequency of consumption of fruits, berries, and vegetables. Our data are drawn from the Norwegian Monitor Survey 2005–2009, a nationally representative and repeated cross section survey. Patterns of inequality are compared by first estimating a multivariate probit model for lifestyles and health, and then by decomposing associated Gini indices and education-related and income-related CIs. We use identical regressors across all six lifestyle and health equations of the multivariate probit model, as well as in decompositions of the associated inequality indices. This specification is different from Contoyannis and Jones (2004) and Balia and Jones (2008), who utilized British panel data to estimate recursive systems in which lifestyles affected future SAH and mortality.[2] The cross sectional nature of our data limits our ability to estimate such dynamic models of health production. However, our main interest is to directly compare important lifestyle and health variables with respect to their correlates and sources of inequality, and not in assessing the actual impact of different lifestyles on health.

As determinants and correlates of lifestyles and health, and as potential sources of inequality in these variables, we consider demographic factors, income, education, occupation, psychological traits, and childhood circumstances. Evidence on the importance of psychological traits such as time preferences, risk aversion and self-control in affecting lifestyles and health is accumulating. As discussed in Balia and Jones (2008), people tend to

[2] In the recursive multivariate probit system of Contoyannis and Jones (2004), lifestyles in 1984 affected SAH in 1991. In Balia and Jones (2008), lifestyles in 1993/94 affected SAH in 1993/94, while lifestyles and SAH in 1993/94 in turn affected mortality in 2003.

invest more in healthy lifestyles when the individual rate of time preference is low. Furthermore, lack of self-control may result in unhealthy lifestyles and thus in poor health (O'Donoghue and Rabin, 2006). Increasing attention is also being paid to the importance of childhood circumstances in affecting adult lifestyles and health. Special emphasis has been placed on variables such as fetal nutrition, social support, and parental education. Recent empirical evidence from the US, Britain and France regarding some of these issues is provided in Case *et al.* (2005), Rosa Dias (2009, 2010) and Trannoy *et al.* (2010). It should be of interest to add empirical evidence regarding these issues in Norway, which has a low level of income inequality and a well-funded welfare state which specifically seeks to avoid that childhood health and learning, later educational and career decisions, and access to health care are to be determined by family background or adult socioeconomic status. Nevertheless, evidence from a number of studies suggests that health and socioeconomic status are closely related also in Norway (Norwegian Directorate of Health, 2005).

2. Data and variables

The Norwegian Monitor Survey is a nationally representative and repeated cross section survey of adults aged 15–95 years. The survey has been conducted biannually since 1985. However, some of our key variables are based on survey questions that were introduced in the 2005 survey, and thus only data from the surveys in 2005, 2007 and 2009 are used.[3] Our sample is further restricted to only include respondents aged 25–74 years, as we want to study individuals who can be expected to having completed most of their education and started earning incomes, and since the sample includes relatively few respondents in the age range 75–95 years. After deleting observations with missing information on any relevant variables (1,995 observations), our final sample consists of 7,738 observations.

[3] The questions that were introduced in the 2005 survey relate to self-reported height and body weight, and parental education.

Each respondent answers an extensive list of questions. The questions related to the selected lifestyle variables and SAH are based on various types of categorical scales. For example, the respondents are asked to indicate their frequency of eating (i) fruits and berries and (ii) vegetables on a ten-point scale ranging from 'never/less than once per month' to 'four times per day'. Similarly, physical activity has an eight-point frequency scale ranging from 'never' to 'once or more per day'. Yet other frequency scales are being used for smoking and fish consumption. SAH is based on the typical five-point scale ranging from 'very bad' to 'very good' health. This use of different categorical scales complicates our intention to compare patterns of inequality in these variables and also body mass. Therefore, we choose to dichotomize each of our six lifestyle and health variables. For robustness purposes, our later analyses have also been conducted using alternative definitions for these variables, as discussed in Section 3.3.

Variable definitions, sample means and associated standard deviations are presented in Table 1. About 59% of the sample exercise at least twice per week, 40% eat fruits and vegetables at least twice per day, 79% eat fish for dinner at least once a week, and 24% are daily smokers. About twelve percent report themselves as obese and 71% report their health status as being good or very good.[4]

Education is categorized into four groups, from having completed only secondary school or less, to having obtained a college or university degree. We have created one dummy variable for each of these four educational classes. About 12% have not completed high school, while 35% have a college or university degree. Income is also divided into four classes using dummy variables. The original survey question on household income included nine response alternatives, each representing a specific income interval. Before dividing

[4] The usual disclaimer applies with respect to strengths and limitations of using SAH and self-reported obesity status as indicators of health, including the possibility of respondents under-reporting their weight and over-reporting their height (Connor Gorber *et al.*, 2007). However, both SAH and obesity have been shown to be closely related to objective measures of health, such as chronic diseases, mental disorders, and mortality (Idler and Benyamini, 1997; Mokdad *et al.*, 2003; Scott *et al.*, 2008).

Table 1
Variable descriptions and summary statistics.

Variable	Description	Mean	S.D.
Lifestyles			
PA	Do physical activity at least twice per week: 1	0.587	0.492
FV	Eat fruits, berries and vegetables at least twice per day: 1	0.399	0.490
FISH	Eat fish for dinner at least once per week: 1	0.788	0.409
NSMOKE	Not smoking cigarettes daily: 1	0.763	0.425
Health			
NOBESE	Body mass index (BMI) (weight in kg/height in meter^2) < 30: 1	0.876	0.330
SAH	Self-assessed health 'good' or 'very good': 1	0.707	0.455
Demographics	(Ref. categories are 'Age 25-34' and 'Living as married')		
Age 35-44	Age 35-44: 1	0.235	0.424
Age 45-54	Age 45-54: 1	0.230	0.421
Age 55-64	Age 55-64: 1	0.251	0.434
Age 65-74	Age 65-74: 1	0.144	0.351
Female	Female: 1	0.535	0.499
Household has children	If any children is living in household: 1	0.454	0.498
Widow	If widowed: 1	0.040	0.196
Divorced	If divorced: 1	0.091	0.288
Single	If single: 1	0.102	0.303
Socioeconomic status	(Ref. cats. are 'Secondary school' and 'Income quartile 1')		
High school	If highest education is high school: 1	0.332	0.471
Some college/univ.	If highest education is some college/university: 1	0.197	0.398
College/univ. degree	If highest education is college/university with degree: 1	0.347	0.476
Income quartile 2	If household income in second quartile: 1	0.261	0.439
Income quartile 3	If household income in third quartile: 1	0.227	0.419
Income quartile 4	If household income in fourth quartile: 1	0.259	0.438
Occupation	(Ref. category is 'Non-manual worker')		
Skilled manual	If skilled manual worker: 1	0.183	0.387
Unskilled manual	If unskilled manual worker: 1	0.066	0.249
On social security/benefit	If on social security or disability benefit: 1	0.101	0.302
Other occupation	If unemployed, student, homemaker, retired, or other: 1	0.263	0.440
Psychological traits			
Pay in installments	Like to pay in instalments: 1[a)]	0.151	0.358
Life insurance	Household has purchased life insurance: 1	0.473	0.499
Self-control	Feel self-control over life outcomes: 1[b)]	0.844	0.363
Childhood circumstances	(Ref. cats. are 'Poor childhood' and 'Lower parental education')		
Childhood ec. average	If family's economic situation normal when 10-15 years: 1	0.654	0.476
Childhood ec. rich	If family well-endowed when 10-15 years: 1	0.129	0.336
Mother high school	If mother's highest education level high school: 1	0.182	0.386
Mother college/univ.	If mother's highest education level college/university: 1	0.125	0.330
Father high school	If fathers's highest education level high school: 1	0.200	0.400
Father college/univ.	If fathers's highest education level college/university: 1	0.199	0.399

Notes: Data pooled from survey years 2005, 2007 and 2009, in total 7,738 individual observations. [a)] Respondent 'partly agrees' or 'totally agrees' in that he/she likes to purchase in instalments. [b)] Respondent 'partly disagrees' or 'totally disagrees' in the statement 'It is of little use to plan for the future, since what happens in life is mostly a matter of being lucky or unlucky anyway'.

income into four quartiles, we (i) set household income to the mid-point value of each income interval, (ii) adjusted for inflation over the survey period 2005–2009, and (iii) adjusted for household size by dividing the resulting income measure by the square root of household size (OECD, 2008).

Some of the variables in Table 1 representing psychological traits and childhood conditions are crudely measured, and the estimated effects of these variables must be interpreted with some caution. However, the use of preference for paying in installments and the purchase of life insurance to proxy time preferences and risk averseness, respectively, is not uncommon (Loewenstein and Prelec, 1992; Cutler *et al.*, 2008). About 15% of the sample indicates a high time preference by liking to pay in installments, 47% indicates to be risk averse by purchasing life insurance, and 15% indicates a lack of self-control by believing that future outcomes mainly depend on being lucky or unlucky. About two-thirds of the sample described their family's economic condition as normal when they were 10–15 years old, and about 13% considered their family to be well-off at that time. Thirteen percent of the participants had mothers with a college or university degree, and twenty percent had fathers with a college or university degree.

3. Empirical methods

As discussed above, a multivariate probit model in combination with decomposition techniques for associated Gini and concentration indices are used to compare patterns of inequality in lifestyles and health.

3.1. Determinants and correlates of lifestyles and health – the multivariate probit model

Our main specification of the multivariate probit model controls for all the variables that are listed in Table 1, plus survey years. Thus, suppressing subscripts for individual *i*, the linear index functions of our multivariate probit system are:

$$
\begin{aligned}
y_{pa^*} &= \boldsymbol{X}_{dem}\boldsymbol{\beta}_{pa} + \boldsymbol{X}_{ses}\boldsymbol{\gamma}_{pa} + \boldsymbol{X}_{occ}\boldsymbol{\delta}_{pa} + \boldsymbol{X}_{psy}\boldsymbol{\pi}_{pa} + \boldsymbol{X}_{chi}\boldsymbol{\lambda}_{pa} + \boldsymbol{X}_{c,y}\boldsymbol{\zeta}_{pa} + \varepsilon_{pa},\\
y_{fv^*} &= \boldsymbol{X}_{dem}\boldsymbol{\beta}_{fv} + \boldsymbol{X}_{ses}\boldsymbol{\gamma}_{fv} + \boldsymbol{X}_{occ}\boldsymbol{\delta}_{fv} + \boldsymbol{X}_{psy}\boldsymbol{\pi}_{fv} + \boldsymbol{X}_{chi}\boldsymbol{\lambda}_{fv} + \boldsymbol{X}_{c,y}\boldsymbol{\zeta}_{fv} + \varepsilon_{fv},\\
y_{fish^*} &= \boldsymbol{X}_{dem}\boldsymbol{\beta}_{fish} + \boldsymbol{X}_{ses}\boldsymbol{\gamma}_{fish} + \boldsymbol{X}_{occ}\boldsymbol{\delta}_{fish} + \boldsymbol{X}_{psy}\boldsymbol{\pi}_{fish} + \boldsymbol{X}_{chi}\boldsymbol{\lambda}_{fish} + \boldsymbol{X}_{c,y}\boldsymbol{\zeta}_{fish} + \varepsilon_{fish},\\
y_{nsmoke^*} &= \boldsymbol{X}_{dem}\boldsymbol{\beta}_{nsmoke} + \boldsymbol{X}_{ses}\boldsymbol{\gamma}_{nsmoke} + \boldsymbol{X}_{occ}\boldsymbol{\delta}_{nsmoke} + \boldsymbol{X}_{psy}\boldsymbol{\pi}_{nsmoke} + \boldsymbol{X}_{chi}\boldsymbol{\lambda}_{nsmoke} + \boldsymbol{X}_{c,y}\boldsymbol{\zeta}_{nsmoke} + \varepsilon_{nsmoke},\\
y_{nobese^*} &= \boldsymbol{X}_{dem}\boldsymbol{\beta}_{nobese} + \boldsymbol{X}_{ses}\boldsymbol{\gamma}_{nobese} + \boldsymbol{X}_{occ}\boldsymbol{\delta}_{nobese} + \boldsymbol{X}_{psy}\boldsymbol{\pi}_{nobese} + \boldsymbol{X}_{chi}\boldsymbol{\lambda}_{nobese} + \boldsymbol{X}_{c,y}\boldsymbol{\zeta}_{nobese} + \varepsilon_{nobese},\\
y_{sah^*} &= \boldsymbol{X}_{dem}\boldsymbol{\beta}_{sah} + \boldsymbol{X}_{ses}\boldsymbol{\gamma}_{sah} + \boldsymbol{X}_{occ}\boldsymbol{\delta}_{sah} + \boldsymbol{X}_{psy}\boldsymbol{\pi}_{sah} + \boldsymbol{X}_{chi}\boldsymbol{\lambda}_{sah} + \boldsymbol{X}_{c,y}\boldsymbol{\zeta}_{sah} + \varepsilon_{sah}, \qquad (1)
\end{aligned}
$$

where PA, FV,…, SAH = 1 if $y_{pa^*}, y_{fv^*},\ldots, y_{sah^*} > 0$ and 0 otherwise, $\boldsymbol{X}_{dem}, \boldsymbol{X}_{ses}, \boldsymbol{X}_{occ}, \boldsymbol{X}_{psy}$ and $\boldsymbol{X}_{chi}$ are vectors of regressors representing the groups of variables in Table 1 for, respectively, demographics, socioeconomic status, occupation, psychological traits, and childhood circumstances. The vector $\boldsymbol{X}_{c,y}$ includes a constant, one, and dummy variables for survey years 2007 and 2009. $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\pi}, \boldsymbol{\lambda}$ and $\boldsymbol{\zeta}$ are corresponding lifestyle-specific and health-specific parameter vectors.

Unlike single equation probit models, this system of probit equations can potentially capture systematic patterns of unobserved individual characteristics, through estimating correlation coefficients ρ between error terms of its different equations, e.g. $\rho_{pa,sah}$= corr(ε_{pa}, ε_{sah}). Thus, controlling for the observed characteristics in Eq. (1), these error correlations indicate to what extent there exist other, unobserved factors, such as for example genetic endowments, which make individuals systematically choose healthy (unhealthy) lifestyles and be in good (poor) health (Balia and Jones, 2008). This is accomplished by modelling, for each respondent, the joint probability of observing his or her particular sequence of responses to our six binary lifestyle and health variables. The error vector $\boldsymbol{\varepsilon}$ is assumed to be distributed multivariate standard normal, $\boldsymbol{\varepsilon} \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Omega})$, with the 6×6 variance-covariance matrix $\boldsymbol{\Omega}$ having values of 1 on its leading diagonal elements, and symmetrical correlation coefficients ρ_{jk} between equations j and k on its off-diagonal elements (Cappellari and Jenkins, 2003). The model is estimated by maximum likelihood. The linear index functions in Eq. (1) enter the log-likelihood function, and our 7,738×6 matrix of lifestyle and health probabilities is simulated using the Geweke–Hajivassiliou–Keane (GHK) simulator.[5] More details on the properties and the technicalities of the multivariate probit model, including advantages of using the GHK simulator, may be found in Cappellari and Jenkins (2003), Contoyannis and Jones (2004) and Balia and Jones (2008).

3.2. Sources of total inequality and socioeconomic inequality in lifestyles and health

The Gini index measures the distribution of a variable y (e.g. SAH) within a population. The closely related concentration index (CI) measures the relationship between y and the distribution of a socioeconomic status indicator (e.g. education). The standard version of the CI (Gini) for y is

$$CI_y = \frac{2}{\mu_y}\operatorname{cov}(y, r), \tag{2}$$

where CI_y is the CI (Gini) calculated for variable y (e.g. SAH), r in the case of CI is the fractional rank of the chosen socioeconomic indicator (e.g. education), and μ_y is the mean of y.[6] The CI has range [-1, 1], where 1 (-1) indicate extreme cases in which all 'good health' is found among those in the absolute highest (lowest) socioeconomic status group (Wagstaff *et al.*, 1991). $Gini_y$ is obtained by replacing r of the socioeconomic status indicator in Eq. (2) by r of y (Wagstaff and van Doorslaer, 2004). The Gini has range [0, 1], where 0 (1) indicate minimum (maximum) levels of total inequality in y. The standard CI (Gini) in Eq. (2) may possess some undesirable properties.[7] As these properties are relevant in this study, we use the following version of the CI (Gini), recently developed by Erreygers (2009a, 2009b):

---

[5] As programming the optimization routine for the multivariate probit model is fairly involved, we used the Stata module *mvprobit* for estimation, which was developed by Cappellari and Jenkins (2003). We follow their recommendation of using about $\sqrt{N}$ draws for the GHK simulator (we use 90 draws).

[6] The fractional rank of a variable x is the integer rank of x divided by the sample size N, i.e. $r_x=i/N$, with i=1, 2,…, N for respondents who have the lowest, the second lowest,…, and the highest recorded values of x, respectively (O'Donnel *et al.*, 2008).

[7] Wagstaff (2005) notes that for binary health variables, the bounds of CI depend on the mean of health and is generally narrower than [-1, 1], while Erreygers (2009a) shows that this generalize to any health variable with a

$$CI_y(E) = \frac{4\mu_y}{b_y - a_y} CI_y = \frac{8}{b_y - a_y} \text{cov}(y, r), \tag{3}$$

where b_y and a_y are the upper and lower limits of y, and CI_y ($Gini_y$) *is* the standard version of the CI (Gini), given by Eq. (2).

Covariances are central to both the CI and the Gini, and these indices may thus be estimated using linear regression.[8] The CI and the Gini may also be decomposed into their contributing factors (Wagstaff *et al.*, 2003). Thus, one might for example estimate the percentage contribution of variables such as age, gender and education to total inequality in SAH. We will decompose the Gini index and the income-related and education-related CI in all our six lifestyle and health variables. The decomposition for $CI_y(E)$ is

$$CI_y(E) = \frac{4\mu_y}{b_y - a_y}\left[\sum_{k=1}^{K} \frac{\hat{\beta}_k \mu_k}{\mu_y} CI_k + \frac{CC_\varepsilon}{\mu_y}\right] = \frac{4}{b_y - a_y}\left[\sum_{k=1}^{K} \hat{\beta}_k \mu_k CI_k + CC_\varepsilon\right], \tag{4}$$

where $\hat{\beta}$ is the estimated coefficient for variable k (e.g. gender) in a linear regression of variables k = 1, 2,…, K on y (e.g. SAH), μ_k is the mean of variable k and CI_k is the standard version of the CI, given by Eq. (2), for variable k with respect to the chosen socioeconomic status indicator (e.g. education). CC_ε is the generalized CI for the error term, which is

finite upper value or a positive lower value. It is therefore generally difficult to compare CIs of variables with different means. Erreygers (2009a, 2009b) further notes that the standard CI (Gini) is not invariant to whether the variable of interest is defined in terms of health or ill health. For example, the CI for a binary indicator for non-smoking will generally not be equal to minus CI for the corresponding indicator for smoking. To account for these and a few other potential shortcomings of the standard CI (Gini), Erreygers (2009a, 2009b) proposes a new version of CI (Gini) − see Eq. (3). In this study, we want to compare CIs (Ginis) across different binary lifestyle and health variables (or more precisely, their predicted linear index functions). Furthermore, in order to present results consistently, we have 'reversed' our unhealthy lifestyle and health variables and made them 'healthy' by defining them as non-smoking and non-obese, as indicated in Table 1. As the standard CI (Gini) (Eq. 2) would be sensitive to differences in means of our lifestyle and health variables and to whether these are defined in terms of health or ill health, we therefore choose to use CI (Gini) as defined in Eq. (3) in this study.

[8] We use regression techniques as described in O'Donnell *et al.* (2008:103) with some adjustments to obtain sample-weighted CIs (Ginis) according to Eq. (3). Also, since our socioeconomic indicators income and education are categorical, we follow the example of e.g. Chen and Roy (2009) by giving equal fractional rank r to ties (their average fractional rank), rather than sort people with equal incomes or education randomly, or by other variables than income or education. For the income-related CIs, we use observed household income instead of the four discrete income quartiles as defined in Table 1 to generate fractional ranks, r. This because observed household income contains more information (i.e., is more 'continuous').

computed as a residual (Balia and Jones, 2008).[9] $Gini_y(E)$ is also decomposed using Eq. (4), but now with CI_k representing the standard CI of variable k with respect to lifestyle or health variable y. Henceforth, we denote $CI_y(E)$ and $Gini_y(E)$ as CI_y and $Gini_y$.

Since all our lifestyle and health variables are binary, we follow the procedure of Balia and Jones (2008) in basing our Gini calculations on predicted probabilities rather than observed outcomes. These individual predictions come from the explained part of the linear index functions of our multivariate probit model, given by Eq. (1).[10] This procedure ensures that we get sufficient variability in the outcome variables for which to calculate the Ginis. However, predicted probabilities are additive in the regressors, and thus only the deterministic part of the decomposition equation can be calculated, i.e. CC_ε in Eq. (4) is not identified (Balia and Jones, 2008). For consistency, we also calculate and decompose the education- and income-related CIs using predicted probabilities, although these in principle could be calculated using observed outcomes, which has been done in robustness tests, as discussed in Section 3.3.

3.3. Inference and robustness

Many and complex associations potentially exist between the different control variables of this study, and between these control variables and the dependent variables for lifestyles and health. For example, time preferences may affect educational decisions, and higher education levels may in turn affect time preferences (van der Pol, 2010). Another example is the complex triangular relationship that exists between income, work status and SAH (Case and Deaton, 2005), which will be discussed in Section 5. The cross sectional nature of our data

[9] Thus, as an example, in order for gender to make an important contribution to the education-related CI in SAH, we must have that (i) the effect of gender on SAH, controlling for other factors (and scaled by the mean of gender), is strong, and (ii) gender and education are strongly correlated, i.e. the education-related CI with respect to gender is large.

[10] Before calculating the Ginis, the linear index functions are transformed to ensure positive predictions. Thus, if $\hat{g}$ denotes the transformed linear index function, we have for example that $\hat{g}_{pa}$=((y_{pa*}-ε_{pa})–min(y_{pa*}-ε_{pa})). This transformation does not affect the percentage contribution of different variables to the Gini in corresponding decomposition analyses (Balia and Jones, 2008: van Doorslaer and Jones, 2003).

makes it particularly difficult to disentangle these and other types of feedbacks, including the identification of causal effects. The results of this study must be viewed in light of these limitations.

Several tests have been conducted to control for robustness of our main results, including analyses that use alternative definitions for our dependent variables. In particular, to utilize more of the information that is available in these variables, we have estimated a multivariate ordered probit model for lifestyles and health (Rosa Dias, 2010), and decomposed associated Gini indices.[11] Also, the education-related and income-related CIs in lifestyles and health have been calculated and decomposed using the observed, binary 0–1 outcomes, rather than predicted probabilities, as described above. In general, our main results are very robust across different model specifications. Due to the interest of space, complete results of these robustness tests will not be presented here, however, a few significant results will be commented on.[12]

4. Correlates of lifestyles and health – results of multivariate probit models

We start comparing patterns of inequality in lifestyles and health by looking at the results of two multivariate probit models. These are reported in Table 2 and Table 3. For ease of interpretation, marginal probabilities are shown instead of original probit parameters.[13] The

[11] In the multivariate ordered probit model and in the associated Gini decompositions, frequency-based measures for physical activity, fruits and vegetables, and fish were used to define these variables in terms of respectively eight, nine, and eight non-monotonically increasing frequency categories. We do not have information on the number of cigarettes smoked daily, and non-smoking was therefore defined as a binary variable, as in the multivariate probit model. Body mass was divided into four groups: (i) BMI < 25.0 (48.6% of the sample), (ii) 25.0 ≤ BMI < 27.5 (24.9%), (iii) 27.5 ≤ BMI < 30.0 (14.2%), and (iv) BMI ≥ 30.0 (12.4%). Group (i) include 1.1% of the respondents that were underweight (BMI < 18.5), while group (iv) include 2.7% of the respondents that had BMI ≥ 35.0, i.e. obese class II or III. SAH was defined using the usual five-point scale ranging from 'very bad' to 'very good'. The joint multivariate ordered probit/binary probit model was estimated using the Stata module *cmp* (Roodman, 2009). This module allows for estimating multivariate normal models that have dependent variables with different scales (e.g., continuous, binary, and ordered variables).
[12] Results of the multivariate ordered probit model and of alternative Gini and CI decompositions are available upon request.
[13] Full results of the multivariate probit models in Table 2 and Table 3, including original parameter estimates and their robust standard errors, are available upon request. The marginal probabilities in Table 2 and Table 3 represent average partial effects. Thus, the marginal probability of each regressor has been calculated for each

results in Table 2 are based on a basic model specification in which we only control for a standard set of demographic and socioeconomic status variables, as defined in Table 1, and survey years. The main purpose of this basic model specification is to assess how education and income are related to lifestyles and health while controlling for a standard set of variables. The results of the full model specification are reported in Table 3. In this specification, we also control for occupational status, psychological traits, and childhood circumstances. The main results in these models are as follows.

Table 2
Multivariate probit model for lifestyles and health – basic specification, marginal probabilities.

	(1) PA	(2) FV	(3) FISH	(4) NSMOKE	(5) NOBESE	(6) SAH
High school	*0.034*	0.026	0.007	0.013	-0.012	0.025
Some college/univ.	**0.101**	**0.090**	**0.061**	**0.104**	-0.006	**0.073**
College/univ. degree	**0.129**	**0.144**	**0.095**	**0.170**	***0.034***	**0.103**
Income quartile 2	**0.045**	0.018	-0.002	*0.023*	*0.021*	**0.102**
Income quartile 3	**0.064**	0.028	*0.028*	**0.049**	**0.044**	**0.166**
Income quartile 4	**0.098**	**0.069**	0.015	**0.054**	**0.051**	**0.195**
Number of obs.						7738
Log-likelihood						-24516.24

Notes: The table shows selected marginal probabilities from a multivariate probit model which also controls for survey years and the group of demographics variables listed under the heading 'Demographics' in Table 1, i.e. variables for age, gender, kids in the household and marital status. The model was estimated using the Stata module *mvprobit* (with 90 draws). Sample weights were applied. Marginal probabilities in **bold**, ***bold italics*** and *italics* are statistically significant at the 99%, 95% and 90% levels, respectively. The marginal probabilities represent average partial effects (see Footnote 13 for details). See Table 1 for the definitions of relevant reference categories.

First, controlling only for basic demographic and socioeconomic variables, clear income and education gradients exist in most lifestyle and health variables.[14] However, some notable exceptions exist. For the consumption of fish, the marginal probabilities of higher income are unclear, and for the consumption of fruits and vegetables, the marginal probability

individual in the sample. The sample means of these individual level calculations represent the average partial effects. Standard errors of these average partial effects are obtained by combining the original probit parameters and their robust standard errors, using the delta method.

[14] Although the marginal probabilities for healthy lifestyles and good health generally increase step-wise with higher education and income levels (i.e., gradients are present), the marginal probabilities for the second lowest education group (High school) and the second lowest income group (Income quartile 2) are in several cases imprecisely estimated.

Table 3
Multivariate probit model for lifestyles and health – full specification, marginal probabilities.

	(1) PA	(2) FV	(3) FISH	(4) NSMOKE	(5) NOBESE	(6) SAH
Age 35-44	0.002	***0.052***	**0.067**	***-0.051***	***-0.041***	**-0.074**
Age 45-54	**0.002**	**0.078**	**0.119**	**-0.086**	-0.015	**-0.082**
Age 55-64	*0.045*	**0.112**	**0.186**	-0.033	0.002	**-0.135**
Age 65-74	**0.088**	**0.149**	**0.205**	0.034	0.016	**-0.111**
Female	**0.068**	**0.183**	**0.044**	**-0.029**	0.014	0.011
Household has children	*-0.031*	-0.015	**0.044**	0.009	0.007	***0.033***
Widow	0.037	*-0.053*	-0.052	-0.017	0.033	*0.043*
Divorced	***0.049***	-0.028	**-0.087**	**-0.098**	0.022	0.010
Single	***0.055***	**-0.060**	**-0.082**	***-0.042***	-0.026	0.001
High school	0.029	0.014	0.001	-0.005	-0.020	0.000
Some college/univ.	**0.089**	***0.056***	**0.050**	**0.072**	-0.022	0.030
College/univ. degree	**0.113**	**0.096**	**0.079**	**0.127**	0.012	***0.045***
Income quartile 2	**0.048**	0.011	-0.002	0.011	0.015	**0.061**
Income quartile 3	**0.065**	0.014	0.026	*0.029*	**0.037**	**0.113**
Income quartile 4	**0.099**	***0.046***	0.011	0.025	**0.040**	**0.123**
Skilled manual	*0.031*	-0.020	0.002	-0.013	0.017	-0.002
Unskilled manual	-0.015	***-0.063***	-0.037	**-0.099**	-0.004	**-0.074**
On social security/benefit	0.015	*-0.040*	0.006	**-0.090**	**-0.050**	**-0.416**
Other occupation	**0.055**	-0.009	0.014	0.011	0.004	**-0.050**
Pay in installments	-0.018	***-0.034***	**-0.042**	**-0.039**	**-0.073**	-0.010
Life insurance	0.019	-0.003	0.004	0.018	-0.002	***0.023***
Self-control	***0.039***	*0.031*	0.007	**0.065**	0.000	**0.060**
Childhood ec. average	-0.010	***0.034***	0.018	**0.044**	0.008	0.019
Childhood ec. rich	0.000	0.017	0.023	*0.029*	***0.034***	***0.039***
Mother high school	0.001	0.022	-0.006	-0.015	***0.026***	0.015
Mother college/univ.	-0.013	***0.058***	0.021	***0.042***	**0.058**	***0.053***
Father high school	-0.009	0.019	0.014	0.004	0.001	-0.017
Father college/univ.	0.031	***0.041***	0.004	0.009	0.017	0.003
2007	0.002	***0.031***	**0.051**	0.017	**-0.032**	-0.004
2009	0.012	**0.089**	*0.020*	**0.033**	**-0.040**	-0.013
Error correlations						
PA		**0.216**	**0.085**	**0.253**	**0.136**	**0.196**
FV			**0.181**	**0.213**	-0.025	***0.050***
FISH				**0.073**	0.044	***0.058***
NSMOKE					**-0.113**	**0.150**
NOBESE						**0.296**
Number of obs.						7738
Log-likelihood						-24158.96

Notes: This multivariate probit model was estimated using the Stata module *mvprobit* (with 90 draws). Sample weights were applied. Marginal probabilities in **bold**, ***bold italics*** and *italics* are statistically significant at the 99%, 95% and 90% levels, respectively. The marginal probabilities represent average partial effects (see Footnote 13 for details). See Table 1 for the definitions of relevant reference categories.

of higher income is significant only in the top quartile. These results suggest that the consumption of healthy foods is not mainly constrained by income. For non-obesity, the positive effect of higher education becomes clear only at the top level, i.e., among those with a college or university degree. This result is somewhat surprising, but several studies have indicated that socioeconomic gradients in body mass may be less clear than corresponding gradients in general health status and various chronic diseases (Sassi, 2010).

Second, the marginal probabilities of higher education levels and higher income quartiles are in most cases substantially reduced when we move from the basic model specification, in Table 2, to the full model specification, in Table 3. For example, in our four lifestyle variables, the marginal probability of having a college or university degree are on average reduced by 22.1% after adding controls for occupational status, psychological traits and childhood circumstances, while the marginal probability of belonging to the top income quartile are on average reduced by 27.6%. In the health variables these average reductions are 61.1% and 29.2%, respectively.[15] Thus, while adding variables for occupational status, psychological traits and childhood circumstances generally reduces the direct influence of current socioeconomic status on current lifestyles and health, these reductions seem to be particularly pronounced in the case of education and health.

Third, income and education are measured on different scales, and it is therefore difficult to assess the importance of income relative to education in one single outcome

[15] Analogously, in a few cases, the marginal probabilities of higher education and income turn from being statistically significant in the basic model specification, in Table 2, to being statistically insignificant in the full model specification, in Table 3. These cases include (i) the marginal probabilities of Income quartile 3 and 4 on non-smoking, (ii) the marginal probability of having a college or university degree on non-obese, and (iii) the marginal probability of having attended college or university (without earning a degree) on SAH. In our alternative multivariate ordered probit specification, where body mass and SAH are defined using multiple discrete categories (see Section 3.3), the effects in (ii) and (iii) remain statistically significant also in the full model specification.

variable. However, it seems that while education is relatively more important than income in predicting healthy lifestyles, the opposite seem to be the case for health.[16]

Fourth, for the role of occupational status, one result in Table 3 clearly stands out; individuals on social security or disability benefit are 41.6 percentage points less likely than others to report being in good or very good health. Being on social security or disability benefit is also significantly associated with the consumption of fruits and vegetables, non-smoking, and non-obesity. Another vulnerable occupation group is unskilled manual workers. Belonging to this group is negatively associated with all indicators of healthy lifestyles and good health, with significant associations found for fruits and vegetables, non-smoking, and SAH.

Fifth, our variables for psychological traits are in several cases significantly related to lifestyles and health. However, no clear pattern seems to stand out. Psychological traits are not relatively more important in predicting lifestyle choices than health outcomes, or vice versa. The respondent's rate of time preference, as proxied by his or her willingness to pay in installments, is significantly associated with all lifestyle variables except physical activity, as well as non-obesity. The variable for sense of self-control over life outcomes is significantly associated with all lifestyle variables except the consumption of fish, as well as SAH. The strongest effect of self-control is found in the case of non-smoking, which is not surprising, given the addictive nature of nicotine. Finally, risk averseness, as proxied by the purchase of life insurance, is significantly associated only with SAH.

Sixth, as for psychological traits, no clear pattern stands out with respect to the role of childhood circumstances. We are unable to identify systematic differences between lifestyle

[16] With one exception, this pattern is also found in our alternative multivariate ordered probit specifications (see Section 3.3). The exception relates to the alternative definition for healthy body mass (four discrete BMI categories). In the multivariate ordered probit models, having a college degree is much more strongly associated with body mass than are the three income dummies. Also, in the full multivariate ordered probit specification (the equivalent to Table 3), having a college degree is the only variable among our six education and income dummies that is significantly associated with the four-category BMI variable. In the basic specification (the equivalent to Table 2), Income quartile 3 and 4 are also significantly associated with the four-category BMI variable (at the 95% level).

and health variables with respect to, for example, the impact of parental education. Being raised by a mother who had completed some form of college or university education is the most important factor among this group of variables. This variable is statistically significant in two of the lifestyle equations and in both health equations, with marginal probabilities in the range 0.042–0.058.[17] It is not surprising that maternal education is more important than paternal education in predicting healthy lifestyles and good health. During the childhood of the majority of our sampled individuals, maternal university education was less common than paternal university education, and thus the former represented a stronger indicator of socioeconomic status. Also, mothers are generally more important than fathers in the process of raising children. Their diet, drinking, and smoking habits during pregnancy may directly affect the child's health later in life (Barker, 1997). Their influence continues through the formation of the child's own lifestyle habits, which are typically predictive of his or her lifestyle habits later in life.

Finally, the residual error terms reported in Table 3 are in several cases strongly correlated. Thirteen out of 15 cross-equation correlation coefficients are statistically significant at the 95% level, and seven out of these correlations are higher than 0.150. Thus, controlling for all the regressors in Table 3, there tend to exist other, unobserved characteristics which make individuals systematically choose healthy (unhealthy) lifestyles and be in good (poor) health. The strongest correlation of unobservable characteristics is found between our two health indicators (0.296). The error correlations are also strong between physical activity, the consumption of fruits and vegetables and non-smoking (0.213–0.253), the consumption of fruits and vegetables and the consumption of fish (0.181), physical activity and SAH (0.196), and non-smoking and SAH (0.150).

[17] For our two health variables, and in particular non-obesity, the education of the respondent's mother seems to be more important than his or her own education in predicting good health. However, in this study we are not able to fully isolate and disentangle the separate effects of parental and own education, as discussed in Section 3.3.

As reported in Rosa Dias (2010), unobserved characteristics in different health affecting lifestyles seem to be more strongly correlated with unobserved characteristics in SAH than other health indicators. While we find less consistent patterns in unobserved characteristics between lifestyles and non-obesity than between lifestyles and SAH, Rosa Dias (2010) found less consistent patterns of unobserved characteristics between lifestyles and long-standing illness, and between lifestyles and mental illness, than between lifestyles and SAH.[18] The strong associations found between SAH and key health affecting lifestyles in terms of common, unobserved individual characteristics provide additional evidence on the relevance of using SAH in empirical work, as it represents not only a good indicator of objective health (Idler and Benyamini, 1997), but also of important production factors of health, here represented by lifestyle choices.

5. Decomposing total inequality and socioeconomic inequality in lifestyles and health

We turn next to inequality in lifestyles and health as measured by Gini and concentration indices. The row 'Gini/CI_{ed}/CI_{inc}' in Table 4 reports the actual index estimates, calculated according to Eq. (3), for respectively total inequality, education-related inequality and income-related inequality in each of our six lifestyle and health variables. The remaining rows in Table 4 report results of the corresponding decomposition analyses, that is, these rows

[18] Although healthy lifestyles are less consistently associated with non-obesity than with SAH in terms of unobserved individual characteristics, it may be argued that these associations are reasonable in light of energy balances. Being a good indicator of energy expenditure, physical activity is clearly positively correlated with non-obesity in terms of unobserved characteristics. The consumption of fruits and vegetables and the consumption of fish are not significantly associated with non-obesity, again in terms of unobserved characteristics. While these diet choices clearly represent indicators of healthy behavior, they are at the same time direct measures of energy intake, and this may explain their ambiguous relation with non-obesity. For non-smoking, there is evidence to suggest that the opposite behavior, i.e. smoking, is associated with lower body weight through affecting ones appetite and metabolic rate (Chiolero *et al.*, 2008). Thus, the negative and significant coefficient for the error correlation between non-smoking and non-obesity in Table 3 (-0.113) is not a surprising result.

Table 4
Decomposing Gini indices and education- and income-related CIs in lifestyles and health – percentage contributions to explained inequality.

	PA			FV			FISH			NSMOKE			NOBESE			SAH		
	Gini	CI_{ed}	CI_{inc}	Gini	CI_{ed}	CI_{inc}	Gini	CI_{ed}	CI_{inc}	Gini	CI_{ed}	CI_{inc}	Gini	CI_{ed}	CI_{inc}	Gini	CI_{ed}	CI_{inc}
Demographics	**37.0**	**-15.9**	**-15.0**	**58.9**	**-15.1**	**-9.8**	**76.1**	**-56.1**	**5.0**	**18.5**	**-1.4**	**-0.4**	**13.2**	**-10.8**	**-12.6**	**18.1**	**10.7**	**-6.9**
Age	12.7	-12.0	-0.8	8.2	-16.7	-2.0	64.8	-68.6	-6.6	9.8	-2.5	-14.0	8.1	-8.3	-4.1	15.5	9.7	-2.5
Female	14.0	0.6	-7.0	47.8	1.5	-20.4	4.3	0.7	-8.7	1.7	-0.2	2.9	1.7	0.2	-2.7	-0.1	0.1	-0.6
Household has children	6.1	-3.2	3.5	0.6	-1.4	1.8	-2.9	8.5	-9.6	0.2	0.6	-0.9	-0.1	1.5	-1.5	3.8	2.3	-1.9
Marital status	4.2	-1.3	-10.6	2.4	1.5	10.8	9.9	3.3	29.9	6.8	0.7	11.7	3.6	-4.2	-4.4	-1.2	-1.3	-1.8
Education	**27.1**	**82.9**	**33.2**	**12.7**	**65.3**	**31.3**	**10.3**	**124.3**	**51.5**	**41.2**	**71.2**	**43.2**	**11.6**	**37.5**	**12.9**	**7.7**	**26.1**	**8.1**
Income	**18.1**	**22.1**	**79.4**	**3.9**	**8.4**	**38.1**	**1.6**	**7.6**	**29.3**	**3.7**	**4.1**	**22.6**	**11.6**	**20.5**	**69.5**	**21.1**	**20.8**	**58.6**
Occupation	**7.9**	**-2.3**	**-10.2**	**3.6**	**10.1**	**11.9**	**3.8**	**5.2**	**-1.9**	**14.5**	**10.5**	**14.4**	**8.1**	**7.2**	**10.2**	**37.7**	**26.3**	**31.5**
Skilled manual	-0.5	-0.8	-0.4	0.6	0.5	0.3	-0.2	-0.1	-0.1	0.3	0.2	0.2	0.9	-0.9	-0.5	-0.1	0.0	0.0
Unskilled manual	1.2	1.6	0.9	2.9	6.4	4.1	2.3	7.1	4.0	6.7	6.2	5.1	0.2	0.8	0.4	2.2	5.2	2.2
On social security/benefit	-0.4	-1.3	-2.0	0.6	3.1	5.9	0.1	-1.0	-1.6	6.6	4.4	10.8	6.7	7.6	11.5	31.9	20.0	25.2
Other occupation	7.6	-1.9	-8.6	-0.4	0.3	1.6	1.6	-0.9	-4.2	0.9	-0.2	-1.7	0.3	-0.3	-1.2	3.7	1.1	4.1
Psychological traits	**5.3**	**7.1**	**7.5**	**3.7**	**4.7**	**5.5**	**3.7**	**5.4**	**6.5**	**10.2**	**6.7**	**10.6**	**17.7**	**5.3**	**5.9**	**6.8**	**6.4**	**5.2**
Pay in installments	1.3	0.8	0.9	2.1	1.3	1.9	3.5	3.2	4.1	2.4	1.0	1.9	17.8	5.6	6.4	0.2	0.3	0.3
Life insurance	0.3	1.3	2.0	0.0	-0.2	-0.3	-0.1	0.5	0.8	1.0	0.8	1.8	-0.1	-0.3	-0.5	2.1	1.1	1.3
Self-control	3.7	5.0	4.5	1.6	3.6	3.9	0.2	1.7	1.6	6.7	4.9	6.9	0.0	0.0	0.0	4.5	5.0	3.6
Childhood circumstances	**3.9**	**5.4**	**3.9**	**7.8**	**21.8**	**14.0**	**0.3**	**13.3**	**7.3**	**9.7**	**7.6**	**6.6**	**32.1**	**44.0**	**20.4**	**8.7**	**10.2**	**4.1**
Childhood ec. situation	0.3	0.1	-0.1	1.2	1.0	1.3	0.4	3.5	2.7	2.7	1.3	1.8	5.7	6.5	3.8	2.1	2.4	1.2
Mother’s education	-0.6	-2.4	-1.1	3.4	11.1	5.7	0.4	7.0	3.4	5.7	4.8	3.4	20.9	28.7	11.4	6.5	7.7	2.6
Father’s education	4.2	7.7	5.1	3.2	9.7	7.0	-0.4	2.8	1.2	1.2	1.5	1.4	5.5	8.8	5.2	0.2	0.1	0.3
Survey year	**0.8**	**0.8**	**1.2**	**9.3**	**4.7**	**9.0**	**4.1**	**0.4**	**2.3**	**2.3**	**1.2**	**3.0**	**5.6**	**-3.8**	**-6.2**	**-0.1**	**-0.6**	**-0.7**
Gini/CI_{ed}/CI_{inc}	.380	.203	.179	.371	.156	.114	.377	.098	.082	.356	.250	.144	.309	.130	.123	.368	.188	.210

Notes: Numbers in bold indicate the percentage contribution of the relevant group of regressors to the explained part of the relevant Gini/CI_{ed}/CI_{inc}. The row ‘Gini/CI_{ed}/CI_{inc}’ reports the actual index estimates, calculated according to Eq. (3).

indicate the percentage contribution of each group of regressors (in bold) to the explained part of the Ginis and the CIs.[20]

The results of the decomposition analyses vary considerably across the three different inequality measures. Not surprisingly, education generally makes a substantial contribution to the education-related CIs in lifestyles and health, with a mean contribution of 67.9%, while income similarly makes a substantial contribution to the income-related CIs, with a mean contribution of 49.6%. However, education and income are much less important in explaining total inequality in lifestyles and health, with average Gini contributions of 18.4% and 10.0%, respectively. For example, gender explains as much as 47.8% of the Gini in fruits and vegetables, but only 1.5% of the corresponding education-related CI. The latter result may be explained by a CI of gender with respect to education that is close to zero.[21] Thus, a study focusing on socioeconomic inequality in the consumption of fruits and vegetables would probably not identify males, but rather those with little education or low income, as the key target group for increased fruits and vegetables consumption. As also other factors than socioeconomic status may be important and relevant elements of inequality in lifestyles and health (Fleurbaey and Schokkaert 2009), we will in the following focus mainly on sources of total inequality in these variables, i.e., on decompositions of the Gini indices in Table 4.

The key contributors to the Gini index are not the same across our different lifestyle and health variables. As indicated above, gender is the key contributor in the consumption of fruits and vegetables. In the consumption of fish, age is clearly most important, with a Gini contribution of 64.8%. This contribution probably reflects generational effects as well as a

---

[20] Full results of the decomposition analyses in Table 4, where contributions to the Ginis and CIs are split into their two main components, i.e., (i) the regresson parameters $\hat{\beta}_k$ of variable *k* in a linear regression of $k = 1, 2, \ldots, K$ on lifestyle or health variable *y* (scaled by the mean value of *k*, μ_k), and (ii) the CI of regressor *k* with respect to education/income (in the case of CI_{ed}/CI_{inc}) or *y* (in the case of Gini), are available upon request.

[21] The education-related CI for Female is 0.0102 (i.e., there are very small gender differences in education). In contrast, the income-related CI for Female is -0.1039 (i.e., higher incomes are concentrated among males). Thus, due to the small education-related CI for Female, gender makes a small contribution to the education-related CI in all six lifestyle and health variables (between -0.2% and 0.7%).

pure age effect.[22] Education is the key factor in explaining differences in physical activity and non-smoking, with Gini contributions of 27.1% and 41.2%. In non-obesity, maternal education is the most important factor, with a Gini contribution of 20.9%, while in SAH being on social security or disability benefit is most important, with a contribution of 31.9%.[23]

Social security status is also important in explaining the income-related CI in SAH, with a contribution of 25.2%. Thus, income and SAH are strongly correlated partly because poor SAH makes individuals exit the labor force prematurely, which in turn affects their incomes negatively due to a shift from earning wages to being on social security. Similar results have been found by Case and Deaton (2005) in the US and van Kippersluis *et al.* (2010) in the Netherlands. However, note that even though we control for the triangular relationship between income, work status and SAH in Table 4, the residual contribution of income to both the Gini and the income-related CI in SAH is still high, at 21.1% and 58.6%. In fact, income contributes more to the Gini in SAH than to the Gini in non-obesity and the four lifestyle variables. Thus, although the effect of poor SAH on income may be largely responsible for the strong income-SAH relationship that is routinely found in empirical work (Smith, 2004), the conventional direction of causality, from income to SAH, also seems to be important.

---

[22]As we in this study use repeated cross section data from the survey years 2005, 2007 and 2009 (i.e., a short time span), we are not able to discriminate between three different time-related dimensions, or effects; age (life course) effects, birth cohort (generational) effects, and period effects (trends). It is difficult to separately identify these three time-related effects even with long-spanning panel data or repeated cross section data, since the third of these variables may always be identified by the other two (e.g., age=survey year-birth year) (Deaton, 1997). Thus, some sort of identifying restrictions must be applied (the typical approach is to ignore cohort effects). With respect to the particular case here of consuming fish, our testing with a longer time-span of the Norwegian Monitor Survey (1987–2009) suggests that there are both strong age effects and birth cohort effects in this diet choice, i.e., older people eat more fish than younger people, and older generations eat more fish than younger generations. Thus, in Table 3 and Table 4 above, the estimated effect of age on consuming fish is probably exaggerated, as this effect also partly reflect generational differences.

[23] When body mass is instead divided into four discrete BMI categories and estimated and decomposed via our alternative multivariate ordered probit model (see Section 3.3), female becomes the clearly most important contributor to the Gini (39.4%), while the contribution of maternal education drops to 9.0%. While the Gini contribution of several variable groups is significantly affected when body mass is instead defined in terms of four discrete BMI categories (which is largely due to the dramatically increased importance of gender), this change is, in addition to maternal education, particularly pronounced for income; while income explains 11.6% of the Gini in non-obesity (Table 4), it explains only 1.1% of the Gini when body mass is defined in terms of four discrete BMI categories.

Education makes a relatively important contribution to the Gini index in all six lifestyle and health variables, although with significant variation, ranging from a low 7.7% in SAH, to a high 41.2% in non-smoking. Income, on the other hand, is relatively unimportant in all lifestyles except physical activity, with Gini contributions of 3.9% in the consumption of fruits and vegetables, 1.6% in the consumption of fish, and 3.7% in non-smoking. Note that the finding in Table 4 of education explaining more of the Gini in lifestyles than the Gini in SAH does not necessarily imply that there are greater educational differences in lifestyles than SAH. Rather, it implies that while both lifestyles and SAH are strongly correlated with education, and while education is strongly correlated with many of the other control variables in Table 3 and Table 4, these other control variables are more directly associated with SAH than with the different lifestyle variables. Thus, the direct contribution of education itself to indices of total inequality and education-related inequality is more 'attenuated' in SAH than in the four lifestyle variables.

Self-control and maternal and paternal education make small but not immaterial contributions to the Gini indices in physical activity, fruits and vegetables, non-smoking and SAH. However, the most notable result with respect to psychological traits and childhood circumstances is the important role of these variables in non-obesity. Preference for paying in installments (17.8%), the economic situation of the respondent's family when he or she was 10–15 years old (5.7%), and maternal (20.9%) and paternal (5.5%) education all make important contributions to the Gini index in non-obesity. Thus, in total, psychological traits and childhood circumstances explain as much as 49.8% of the Gini in non-obesity, which compares against an average contribution of these factors of only 6.0% in our four lifestyle variables and SAH. However, note that these are relative Gini contributions which add up to one hundred percent across the included regressors. Thus, the strong contribution of psychological traits and childhood circumstances to the Gini index in non-obesity partly reflects that the other control groups in Table 4 are unimportant in explaining inequality in this variable. In SAH, most control groups seem to be important in explaining inequality, while varying control groups such as age, gender and own education are very important in explaining inequality in the four lifestyle variables.[24] Thus, in absolute terms, psychological traits and childhood circumstances may be equally important in explaining inequality in our three lifestyle variables and SAH as in non-obesity.

6. Conclusions

The main purpose of this paper has been to compare patterns of inequality in health, represented by self-assessed health (SAH) and obesity, with patterns of inequality in lifestyle choices central to the production of health, represented by physical activity, smoking and diet quality. Using decomposition techniques for the Gini index, we find that education is generally an important source of total inequality in these lifestyle and health variables, while income is mainly important in physical activity and the two health variables. In several cases, education and income are clearly outranked by other factors in terms of explaining total inequality, such as gender in the consumption of fruits and vegetables, age in the consumption of fish and maternal education in non-obesity.

To prevent inequalities in health, policies need to target inequalities in the key production factors of health, including the lifestyle choices considered in this study. Thus, for policy purposes, evidence on patterns of inequality in final health, which is typically what is being collected in empirical work, is mainly relevant to the extent that such patterns are

[24] This hypothesis is partly supported by results of the main alternative model specification in our robustness tests, where body mass is instead defined in terms of four discrete BMI categories (see Section 3.3). With this alternative variable definition for body mass, the total Gini contribution of variables for psychological traits and childhood circumstances drops considerably, to 16.2% (from 49.8% in Table 4). And, this substantial reduction is not mainly the result of psychological traits and childhood circumstances being unimportant in explaining population differences in the four-category BMI variable, but is rather the result of gender being a substantially more important source of inequality in the four-category BMI variable than in the binary variable for being non-obese (as Female explains respectively 39.4% and 1.7% of the Gini indices in these two variables. See also Footnote 22).

representative of patterns of inequality in important, underlying production factors of health. The heterogeneous patterns of inequality found across different lifestyle and health variables in this study suggest that one should be careful in making such assumptions in general. However, our study has several limitations, including its use of repreated cross section data. Thus, additional studies that focus explicitly on mapping patterns of inequality in key factors of health production, in addition to final health itself, are needed.

Population differences in lifestyles and health by socioeconomic status have achieved most of the attention in the literature on health inequalities (van Doorslaer and Van Ourti, 2011). However, as discussed in Fleurbaey and Schokkaert (2009), more attention should probably also be given to other and in some cases perhaps more important sources of inequality, such as age and gender in the examples above, rather than simply label such sources of inequality as 'legitimate' or 'unavoidable'. For example, it is clearly possible to avoid having gender differences in the consumption of fruits and vegetables, and achieving it is a workable policy goal, through for example targeted nutrition campaigns. To efficiently improve overall population health and at the same time reduce the variance of health, one should search for key sources of population differences in single, important production factors of health, including various lifestyle choices, and in turn design tailored policies for each of these factors.

References

Balia S, Jones AM. 2008. Mortality, lifestyle and socio-economic status. *Journal of Health Economics* **27**: 1–26.

Barker DJP. 1997. Maternal nutrition, fetal nutrition and diseases in later life. *Nutrition* **13**: 807–813.

Cappellari L, Jenkins SP. 2003. Multivariate probit regression using simulated maximum likelihood. *Stata Journal* **3**: 278–294.

Case A, Deaton A. 2005. Broken down by work and sex: how our health declines. In *Analyses in the economics of aging.* Wise DA (Ed.). Chicago University Press: Chicago.

Case A, Fertig A, Paxson C. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics* **24**: 365–389.

Chen Z, Roy K. 2009. Calculating concentration index with repetitive values of indicators of economic welfare. *Journal of Health Economics* **28**: 169–175.

Chiolero A, Faeh D, Paccaud F, Cornuz J. 2008. Consequences of smoking for body weight, body fat distribution, and insulin resistance. *American Journal of Clinical Nutrition* **87**: 801–809.

Connor Gorber S, Tremblay M, Moher D, Gorber B. 2007. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity Reviews* **8**: 307–326.

Contoyannis P, Jones AM. 2004. Socio-economic status, health and lifestyle. *Journal of Health Economics* **23**: 965–995.

Costa-Font J, Gil J. 2008. What lies behind socio-economic inequalities in obesity in Spain? A decomposition approach. *Food Policy* **33**: 61–73.

CSDH. 2008. Closing the gap in one generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health. World Health Organization: Geneva.

Cutler DM, Finkelstein A, McGarry K. 2008. Preference heterogeneity and insurance markets: explaining a puzzle of insurance. NBER Working Papers 13746, National Bureau of Economic Research, Inc.

Deaton A. 1997. *The analysis of household surveys: a microeconomic approach to development policy (World Bank).* The Johns Hopkins University Press: Baltimore.

van Doorslaer E, Jones AM. 2003. Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics* **22**: 61–87.

van Doorslaer E, Van Ourti T. 2011. Measuring inequality and inequity in health and health care. In *The Oxford handbook of health economics.* Glied S, Smith PC (Eds.). Oxford University Press: Oxford.

Erreygers G. 2009a. Correcting the concentration index. *Journal of Health Economics* **28**: 504–515.

Erreygers G. 2009b. Can a single indicator measure both attainment and shortfall inequality? *Journal of Health Economics* **28**: 893–901.

Fleurbaey M, Schokkaert E. 2009. Unfair inequalities in health and health care. *Journal of Health Economics* **28**: 73–90.

Idler EL, Benyamini Y. 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior* **38**: 21–37.

van Kippersluis H, O'Donnell O, van Doorslaer E, van Ourti T. 2010. Socioeconomic differences in health over the life cycle in an egalitarian country. *Social Science & Medicine* **70**: 428–438.

Loewenstein G, Prelec D. 1992. Anomalies in intertemporal choice: evidence and an interpretation. *Quarterly Journal of Economics* **107:** 573–597.

Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, Marks JS. 2003. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *The Journal of the American Medical Association* **289**: 76–79.

Norwegian Directorate of Health. 2005. The challenge of the gradient. IS-1245: Oslo.

O'Donnell O, van Doorslaer E, Wagstaff E, Lindelow M. 2008. *Analyzing health equity using household survey data*. World Bank Institute: Washington, DC.

O'Donoghue T, Rabin M. 2006. Optimal sin taxes. *Journal of Public Economics* **90**: 1825–1849.

OECD. 2008. *Growing unequal? Income distribution and poverty in OECD countries*. OECD Publishing: Paris.

van der Pol M. 2011. Health, education and time preference. *Health Economics* **20**: 917–929.

Roodman D. 2009. Estimating fully observed recursive mixed-processes models with *cmp*. Working Paper No. 168, Center for Global Development.

Rosa Dias P. 2009. Inequality of opportunity in health: evidence from a UK cohort study. *Health Economics* **18**: 1057–1074.

Rosa Dias P. 2010. Modeling opportunity in health under partial observability of circumstances. *Health Economics* **19**: 252–264.

Sassi F. 2010. *Obesity and the economics of prevention: fit not fat*. OECD Publishing: Paris.

Scott KM, McGee MA, Wells JE, Oakley Browne MA. 2008. Obesity and mental disorders in the adult general population. *Journal of Psychosomatic Research* **64**: 97–105.

Smith, JP. 2004. Unraveling the SES-health connection. *Population and Development Review,* Supplement: Aging, Health, and Public Policy **30**: 108–132.

Trannoy A, Tubeuf S, Jusot F, Devaux M. 2010. Inequality of opportunities in health in France: a first pass. *Health Economics* **19**: 921–938.

Vallejo-Torres L, Morris S. 2010. The contribution of smoking and obesity to income-related inequalities in health in England. *Social Science & Medicine* **71**: 1189–1198.

Wagstaff A, Paci P, van Doorslaer E. 1991. On the measurement of inequalities in health. *Social Science & Medicine* **33**: 545–557.

Wagstaff A, van Doorslaer E, Watanabe N. 2003. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics* **112**: 207–223.

Wagstaff A, van Doorslaer E. 2004. Overall versus socioeconomic health inequality: a measurement framework and two empirical illustrations. *Health Economics* **13**: 297–301.

Wagstaff A. 2005. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics* **14**: 429–432.

Paper 2

Health inequalities over the adult life course: the role of lifestyle choices

Arnstein Øvrum[a,b*], Geir Wæhler Gustavsen[b] and Kyrre Rickertsen[a,b]

[a]Department of Economics and Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway
[b]Norwegian Agricultural Economics Research Institute, P.O. Box 8024 Dep, N-0030 Oslo, Norway

Abstract

The relationship between socioeconomic status and health is dynamic and evolves throughout the adult life course. However, relatively little empirical attention has been directed to the role of health affecting lifestyle choices in explaining these dynamics. Using Norwegian repeated cross-section data from the period 1997–2009, this study explores how the income and education gradients in physical activity, the consumption of fruits and vegetables, cigarette smoking and self-assessed health evolve over the age range 25–79 years. The findings indicate that while the education gradients in physical activity and the consumption of fruits and vegetables remain relatively stable throughout the adult life course, the education gradient in smoking is clearly decreasing in age. Further, with the exception of the income gradient in physical activity among females, the income gradients in lifestyles are generally concave in age and slightly decreasing in older age. However, the role of lifestyles in moderating the relationship between income and self-assessed health appears modest. This result partly reflects that while the income gradients in lifestyles decrease substantially once we control for education, the reverse is not true. Overall, while income and education differences in lifestyles should generally contribute to cumulative advantage effects in health by socioeconomic status over the adult life course, our results provide some evidence of increased health consciousness and associated lifestyle improvements in age

[*] Corresponding author. E-mail: arnstein.ovrum@nilf.no. This paper was submitted to a peer-reviewed journal in November 2011.

among lower socioeconomic status groups. This could potentially contribute to reducing cumulative advantage effects in health by socioeconomic status at older ages.

JEL classification: D12; D91; I12; I14; I18

Keywords: age; inequality; life course; lifestyles; self-assessed health; socioeconomic status

1. Introduction

An increasingly large literature seeks to improve our understanding of why indicators of socioeconomic status and health are so strongly associated. Acknowledging the dynamic nature of health production, this literature has partly focused on how socioeconomic inequalities in health evolve over the adult life course. The current empirical evidence on this important issue is mixed, partly because different indicators of socioeconomic status and health have been investigated (Kim and Durden, 2007). However, two main patterns of results stand out.

In some studies, health differences by socioeconomic status are found to be increasing in age throughout the adult life course (Ross and Wu, 1996; Wilson *et al.*, 2007). These results correspond with the cumulative advantage hypothesis. This hypothesis asserts that throughout the life course, socioeconomic status is closely associated with daily investments in the production of poor and good health. Gradually, these investments result in a relatively more rapid deterioration of health among lower than higher socioeconomic status groups. In contrast, in other studies health differences by socioeconomic status are found to be increasing in age until late midlife (50–60 years of age), after which they level off or begin to decrease (Beckett, 2000; Huijts *et al.*, 2010; van Kippersluis *et al.*, 2010). The results from these studies are then supportive of the cumulative advantage hypothesis until late midlife, but with an age-as-leveler hypothesis thereafter. More particularly, biological factors (arguably somewhat randomly distributed across people of different socioeconomic status) become increasingly important with older age in determining health, thus downplaying the role of socioeconomic status (Herd, 2006). Also other factors may contribute to age-as-leveler effects in health, including sample selection (Kim and Durden, 2007), cohort effects (Lynch, 2003) and labor market characteristics (Case and Deaton, 2005; van Kippersluis *et al.*, 2010). For example, according to the results in Case and Deaton (2005) and van Kippersluis *et al.* (2010), the strong correlation that typically exists between income and self-assessed health during late midlife mostly reflects the effect of poor health on premature exit from the labor force. This in turn negatively affects incomes because of the shift from wage earning to a reliance on social security payments.

While the above factors may be important in explaining why income and education differences in health vary over the adult life course, there has been relatively little empirical attention directed to the role of health affecting lifestyle choices. For example, do the education and income gradients in physical activity, dietary behavior and cigarette smoking remain stable over the adult life course? Alternatively, do they increase, become smaller, or fluctuate? Moreover, are such life course patterns similar across different lifestyles and across education and income? If education and income gradients in lifestyles remain stable (or increase) over the adult life course, we would expect the corresponding gradients in health, all other things being equal, to be gradually increasing in age because of the long-term, cumulative nature of health production. On the other hand, people of lower socioeconomic status may grow more health conscious and thus engage in healthier lifestyles when they reach late midlife and possibly find themselves in a relatively poor state of health, and thus realize that good health investments are important for longevity. If so, this could contribute to age-as-leveler effects in health, at least to the extent that such changes at older ages are relatively larger among people in lower than higher socioeconomic status groups.

To address these concerns, this paper examines how education and income gradients in important lifestyle and health indicators evolve over the adult life course (the period between 25 and 79 years of age). For this purpose, we employ repeated cross-section data from the Norwegian Monitor Survey 1997–2009. We measure health by self-assessed health (SAH), while physical activity (PA), the consumption of fruits and vegetables (FV) and not smoking cigarettes (NSMOKE) represent lifestyles. These lifestyle indicators are closely associated with the risk of major health outcomes, including type II diabetes, cardiovascular disease and certain types of cancer (World Health Organization, 2003). We analyze the association between age, income, education, lifestyles and SAH using regression models. Sensitivity of the age-specific income and education gradients are assessed by the stepwise inclusion of additional control variables in our models, including occupational status and a variety of sociodemographic characteristics.

2. Data and variables

The Norwegian Monitor Survey is a nationally representative and repeated cross-section survey of adults aged 15–95 years. The survey has been conducted every second year since 1985. The question on SAH was not part of the survey before 1997, and thus only data from the period 1997–2009 are used. We only include respondents aged 25–79 years as we wish to study individuals who can be expected to having completed most of their education and started earning incomes. The sample included relatively few respondents in the age range 80–95 years. After deleting observations with missing information on any of the relevant variables, our final sample comprises 21,706 individual observations.

In the survey, each individual responds to an extensive list of questions. The questions related to PA, FV, NSMOKE and SAH are based on various types of categorical scales. The respondents are asked to indicate their frequency of intake for nine types of fruits and vegetables on the following scale; ‘daily’; ‘3−5 times per week’; ‘1−2 times per week’; ‘2−3 times per month’; ‘about once per month’; ‘3−11 times per year’; ‘rarer’; or ‘never’. Similarly, physical activity has an eight-point frequency scale ranging from ‘never’ to ‘once or more per day’. The respondents also answered if they smoked cigarettes ‘daily’, ‘sometimes’ or ‘never’ at the time of the survey, while SAH is based on the typical five-point scale ranging from ‘very poor’ to ‘very good’ health. To facilitate the comparison of education and income gradients over the adult life course, we have chosen to dichotomize each of these categorical variables. Table 1 provides the descriptions and sample means of these and other relevant variables in this study.

Table 1
Variable descriptions and sample means.

Variable	Description	Mean
PA	Undertake physical activity at least twice per week	0.518
FV	Eat fruits, berries and vegetables at least twice per day	0.485
NSMOKE	Not smoking cigarettes daily	0.702
SAH	Self-assessed health is ‘good’ or ‘very good’	0.689
E_1	Lower secondary education (9 years of education) or less	0.168
E_2	Upper secondary education (12 years of education)	0.359
E_3	Has attended some university or college	0.179
E_4	Has obtained a university or college degree	0.295
I_1	Age-group survey-year specific income quartile 1	0.257
I_2	Age-group survey-year specific income quartile 2	0.252
I_3	Age-group survey-year specific income quartile 3	0.248
I_4	Age-group survey-year specific income quartile 4	0.243
A	Respondent age	47.57
Female	Female	0.536
Children	Any children living in household	0.462
(Living as) married	If married or living as married	0.727
Widowed	Widowed	0.047
Divorced	Divorced	0.096
Single	Single	0.130
Non manual	Nonmanual worker	0.382
Skilled manual	Skilled manual worker	0.173
Unskilled manual	Unskilled manual worker	0.076
On social security	On social security or disability benefit	0.088
Other occupations	Unemployed, student, homemaker, retired, or other	0.281

Notes: Variable means using all 21,706 observations. All variables except age (A) are dummy variables taking a value of one if response to description is yes and zero otherwise.

We categorize education into four groups, ranging from having completed only lower secondary education (9 years of education) or less (E_1), to having obtained a university or college degree (E_4). We divide household income into age-group survey-year specific income quartiles (I_1–I_4), with each age group comprising a five-year interval (e.g., people aged 25–29 years). The original survey question on household income included nine response alternatives, each representing a specific income interval. Before dividing income into age-group survey-year specific quartiles, we (i) set household income to the midpoint value of each income interval, and (ii) adjusted for household size by dividing the resulting income measure by the square root of household size (OECD, 2008).

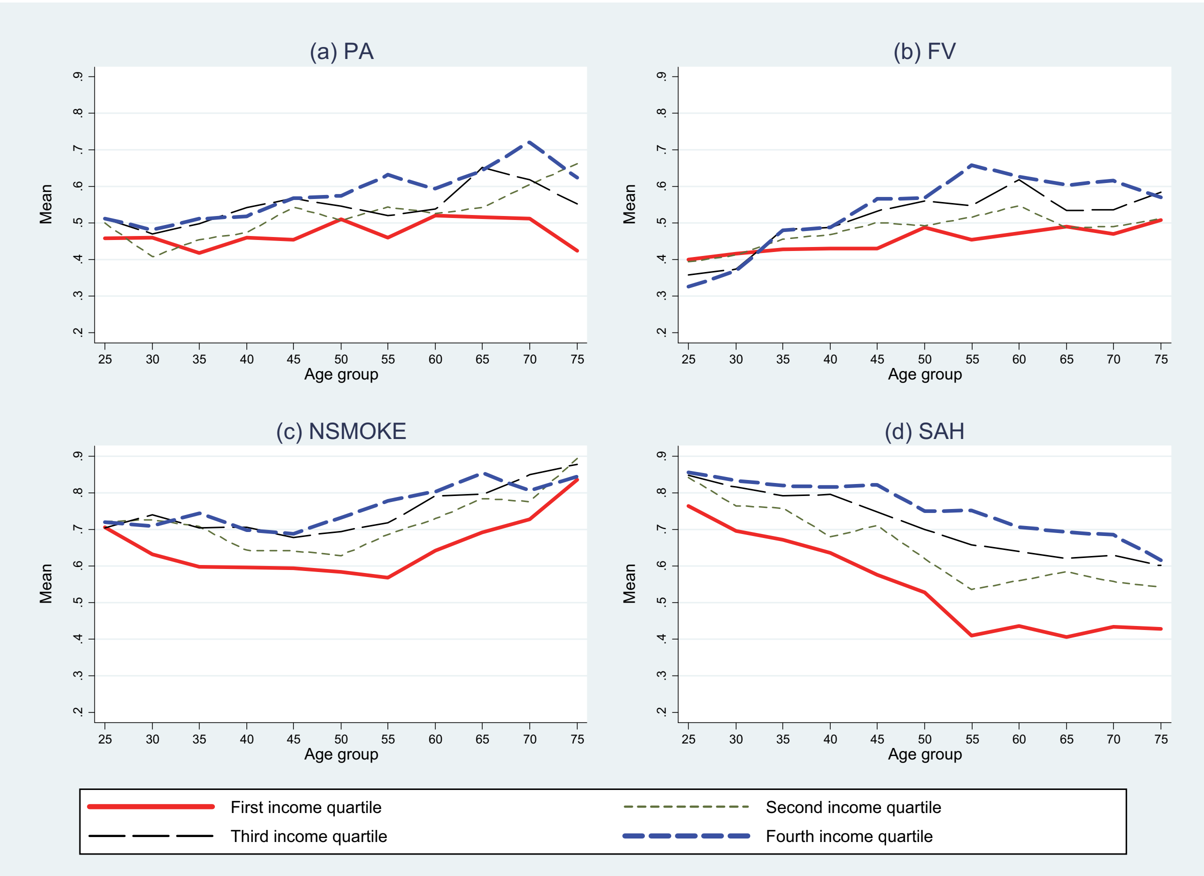


Fig. 1. Mean values for lifestyles and self-assessed health, split by five-year age groups and age-group survey-year specific income quartiles.

Figs. 1 and 2 depict life course variation in lifestyles and SAH by income and education, respectively. These figures essentially illustrate the sample means of PA, FV, NSMOKE and SAH for each income quartile and each education group at each five-year age

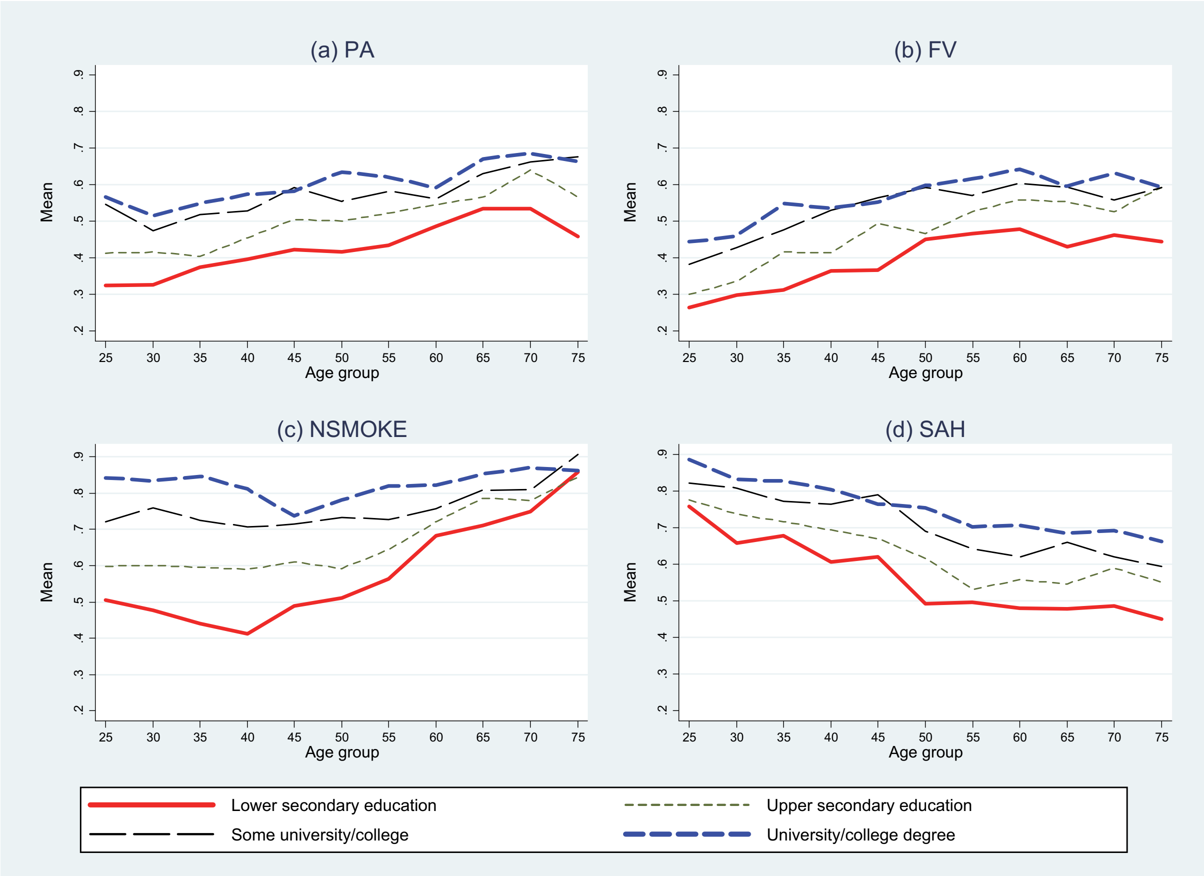


Fig. 2. Mean values for lifestyles and self-assessed health, split by five-year age groups and the four education groups.

interval. As shown, there are generally clear income gradients and particularly clear education gradients in lifestyles and SAH at most stages of the adult life course. The main exceptions are the generally small income gradients in lifestyles at age 25–29 years and the small education and income gradients in NSMOKE at age 75–79 years. Life course variation in the gradients is most evident in the case of income and SAH, with the gradient clearly peaking at age 55–59 years, and in the case of education and NSMOKE, with the gradient clearly declining over the adult life course. However, Figs. 1 and 2 are based on sample means and do not control for confounding factors such as other sociodemographic characteristics and period and cohort effects. We next describe the estimation strategy used to account for these factors.

3. Estimation method

We employ linear probability models (LPM) to predict how the income and education gradients in our three lifestyle variables and SAH evolve over the adult life course. While we obtained very similar results when running logit or probit models as alternatives, LPM coefficients are easier to interpret when using interaction variables (Baum and Ruhm, 2009). For the models focusing on income gradients, our three basic model specifications for lifestyle or health variable y for individual i are:

$$y_i = \alpha+\beta_1 A_i+\beta_2 A_i^2+\beta_3 I_{2i} \qquad +\beta_4 I_{3i} \qquad +\beta_5 I_{4i} \qquad +\mathbf{X}_i'\boldsymbol{\delta} +\varepsilon_i, \quad (1)$$

$$y_i = \alpha+\beta_1 A_i+\beta_2 A_i^2+\beta_3 I_{2i}+\beta_4 A_i\cdot I_{2i} \qquad +\beta_5 I_{3i}+\beta_6 A_i\cdot I_{3i} \qquad +\beta_7 I_{4i}+\beta_8 A_i\cdot I_{4i} \qquad +\mathbf{X}_i'\boldsymbol{\delta} +\varepsilon_i, \quad (2)$$

$$y_i = \alpha+\beta_1 A_i+\beta_2 A_i^2+\beta_3 I_{2i}+\beta_4 A_i\cdot I_{2i}+\beta_5 A_i^2\cdot I_{2i}+\beta_6 I_{3i}+\beta_7 A_i\cdot I_{3i}+\beta_8 A_i^2\cdot I_{3i}+\beta_9 I_{4i}+\beta_{10} A_i\cdot I_{4i}+\beta_{11} A_i^2\cdot I_{4i}+\mathbf{X}_i'\boldsymbol{\delta} +\varepsilon_i, \quad (3)$$

where A_i is the age of individual i centered at age 30, I_2, I_3 and I_4 denote membership of the second, third and fourth income quartiles as defined in Table 1, **X** is a vector of additional control variables, and ε is the stochastic error term. In Model 1 (Eq. 1) the probability of lifestyle or health variable y is explained by a second-degree polynomial in age, indicators of income quartiles, and control dummies. Model 2 (Eq. 2) allows for the marginal effects of higher income quartiles to change linearly in age, while Model 3 (Eq. 3) allows for these marginal effects to change nonlinearly in age. Thus, while Model 2 facilitates, for example, the analysis of cumulative advantage effects in y by income over the adult life course, Model 3 is more flexible in that it facilitates the analysis of cumulative advantage effects followed by age-as-leveler effects at older ages (Beckett, 2000). Comparable models focusing on the education gradients are identical to Eqs. (1)–(3) with the exception that we replace I_2, I_3 and I_4 with dummy variables representing the three highest-level education groups (E_2, E_3 and E_4) as defined in Table 1.

Sensitivity of the income and education gradients will be assessed by varying what variables are included in vector **X** in Eqs. (1)–(3). We denote these different submodels a, b, c and d. All models control for gender and include dummies for the survey years and the five-year birth cohorts. Models with no additional covariates in vector **X** will be denoted as, for example, Model 3a. In Model 3b, the vector **X** also includes education in the models that focus on income gradients and income in the models that focus on education gradients, and dummies for marital status and having children. Model 3c extends Model 3b by controlling for occupational status, including being on social security or being a nonmanual, skilled manual or unskilled manual worker. Finally, Model 3d extends Model 3b by controlling for PA, FV and NSMOKE, i.e., the three lifestyle variables. Model 3d is estimated only for SAH.

In our models, we treat age, period and cohort effects as fixed effects. The linear dependence between respondent age, birth year and survey year is relieved by allowing for nonlinear effects in age and by using five-year birth cohorts (Sarma *et al.*, 2011). We also tested alternative strategies for estimating age, period and cohort effects, including the random intercept model (O'Brien *et al.*, 2008) and the cross-classified model (Reither *et al.*, 2009). The estimated age effects, which are the focus of this study, are very similar across these alternative model specifications.

We also estimated the models separately by gender. For robustness purposes, we also estimated the models using alternative definitions of age, income and education. In this alternative model specification, we replaced the continuous age variables in Eqs. (1)–(3) with five-year age dummies, and the income and education dummies with the logarithm of income and a continuous education variable. We comment on the results of this alternative model specification and the gender specific models where relevant. Finally, the models were re-estimated using alternative variable definitions for PA, FV and SAH (ordered PA and SAH variables and FV in number of intakes per day). The results (not shown) suggest that the main conclusions of the study are not sensitive to how we define the dependent variables in our models.

4. Results

4.1. Income, lifestyles and SAH over the adult life course

Table 2 presents selected parameter estimates from the linear probability models focusing on income gradients in lifestyles and SAH. Table 3 in the next section provides analogous estimates from the models focusing on education gradients. The column headings indicate the different model specifications discussed earlier. Because of space considerations, the tables only detail the parameters for age and income. Further, the estimated age and income effects in PA and FV were largely unaffected after controlling for occupational status, as will be illustrated graphically below, and so we do not provide the results of Model 3c for either of these lifestyle variables.

After controlling for age, gender, survey years and birth cohorts, we can observe clear overall income gradients in the three lifestyle variables as well as SAH (Model 1a). To the extent that these variables are comparable, we can see that the income gradient is steeper in SAH than in underlying, health affecting lifestyles. For example, on average, people in the fourth income quartile are as much as 22.2 percentage points more likely to report being in good or very good health than those in the first income quartile.

Because of interactions between the age and income variables, the parameters of Models 2a–3d in Table 2 are more difficult to interpret than the parameters of Model 1a. To proceed with our analysis, we will mainly focus on graphically comparing patterns of results for the first and the fourth income quartiles. Before turning to this graphical analysis, we note the following main patterns of results from Table 2; (i) there is generally significant life course variation in the income gradients in lifestyles and SAH; (ii) this life course variation is

Table 2
Linear probability models for the association between age, income, lifestyles and SAH.

Model	(1a)	(2a)	(3a)	(3b)	(1a)	(2a)	(3a)	(3b)	(3c)	(3d)
	Physical activity (PA) models				*Nonsmoking (NSMOKE) models*					
A	0.0354	0.0287	***0.0678***	***0.0758***	0.0201	0.0183	–0.0139	–0.0212	–0.0075	
A^2	–0.0045	–0.0046	**–0.0144**	**–0.0151**	0.0045	0.0047	**0.0118**	**0.0162**	**0.0124**	
I_2	**0.0575**	0.0193	***0.0495***	*0.0378*	**0.0683**	**0.0593**	**0.0543**	0.0250	0.0195	
$A\cdot I_2$		**0.0212**	*–0.0450*	*–0.0452*		0.0051	0.0164	0.0126	–0.0008	
$A^2\cdot I_2$			**0.0161**	**0.0156**			–0.0028	–0.0017	0.0016	
I_3	**0.0972**	**0.0942**	**0.1142**	**0.0812**	**0.1096**	**0.0973**	**0.0768**	***0.0422***	***0.0335***	
$A\cdot I_3$		0.0018	*–0.0409*	–0.0361		0.0072	***0.0479***	0.0273	0.0082	
$A^2\cdot I_3$			*0.0103*	0.0086			***–0.0098***	–0.0070	–0.0020	
I_4	**0.1323**	**0.1310**	**0.1602**	**0.1054**	**0.1252**	**0.1579**	**0.1211**	**0.0719**	**0.0592**	
$A\cdot I_4$		0.0010	***–0.0580***	***–0.0568***		***–0.0176***	***0.0516***	0.0224	0.0025	
$A^2\cdot I_4$			***0.0142***	***0.0143***			**–0.0165**	***–0.0121***	–0.0069	
R^2	0.0290	0.0297	0.0305	0.0409	0.0417	0.0426	0.0434	0.0790	0.0848	
	Fruits and vegetables (FV) models				*Self-assessed health (SAH) models*					
A	**0.1430**	**0.1348**	**0.1034**	**0.0888**	–0.0375	*–0.0472*	***–0.0648***	**–0.0774**	–0.0190	**–0.0838**
A^2	**–0.0221**	**–0.0221**	**–0.0147**	*–0.0096*	–0.0016	–0.0017	0.0019	0.0057	–0.0072	0.0058
I_2	**0.0430**	0.0277	0.0135	–0.0125	**0.1095**	**0.0885**	**0.0898**	**0.0763**	**0.0594**	**0.0709**
$A\cdot I_2$		0.0083	*0.0396*	*0.0395*		0.0114	0.0086	0.0105	*–0.0380*	0.0129
$A^2\cdot I_2$			–0.0076	–0.0071			0.0007	–0.0001	***0.0112***	–0.0012
I_3	**0.0913**	**0.0572**	*0.0393*	0.0204	**0.1728**	**0.1360**	**0.1303**	**0.1189**	**0.1016**	**0.1074**
$A\cdot I_3$		***0.0184***	***0.0550***	0.0356		**0.0199**	0.0299	0.0225	***–0.0422***	0.0227
$A^2\cdot I_3$			–0.0088	–0.0060			–0.0023	–0.0021	***0.0123***	–0.0022
I_4	**0.1199**	**0.0797**	***0.0495***	0.0251	**0.2220**	**0.1745**	**0.1459**	**0.1316**	**0.1115**	**0.1152**
$A\cdot I_4$		**0.0216**	**0.0799**	***0.0534***		**0.0255**	**0.0782**	**0.0667**	–0.0033	**0.0690**
$A^2\cdot I_4$			***–0.0140***	*–0.0101*			***–0.0126***	***–0.0118***	0.0036	***–0.0119***
R^2	0.0836	0.0842	0.0846	0.1017	0.0753	0.0762	0.0767	0.0873	0.1310	0.1065

Notes: All models control for gender, survey years and birth cohorts. Models 3b–3d also control for education, marital status and having children. In addition, Model 3c controls for occupational status, while Model 3d controls for PA, FV and NSMOKE. A denotes age (centered at 30 years of age) and I_2, I_3 and I_4 denote age-group survey-year specific income quartiles 2, 3 and 4, respectively (the reference group is income quartile 1 (I_1)). See Table 1 for further variable definitions. Parameters involving A and A^2 are multiplied by 10 and 10^2, respectively. Parameters in **bold**, ***bold italics*** and *italics* are statistically significant at 99%, 95%, and 90% levels using robust standard errors, respectively. Sample weights are applied. All models are based on 21,706 observations.

usually nonlinear (Model 3a); and (iii) the income gradients are in some cases quite sensitive to the addition of extra control variables to the models (Models 3b–3d).

Based on the results of Model 3a in Table 2, Fig. 3 shows the predicted age trajectories in PA, FV, NSMOKE and SAH for the first and fourth income quartiles, as well as the absolute differences in predicted probabilities between these quartiles. We refer to these differences as the income gradient. These and later predictions are calculated at the mean values of the additional covariates (**X**) that are included in the models.

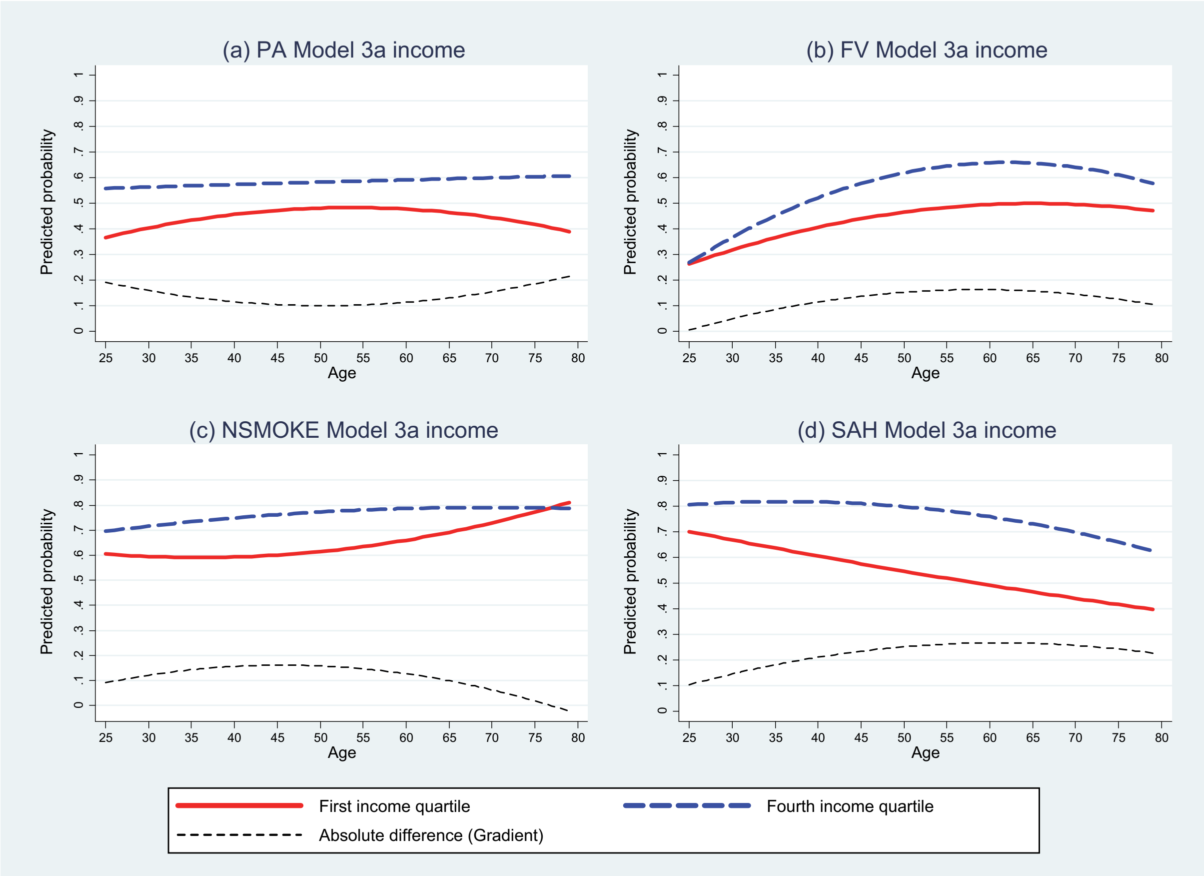


Fig. 3. Predicted age trajectories in lifestyles and self-assessed health for people in the first and fourth income quartiles. Predictions based on the results of Model 3a in Table 2 and calculated at the mean values of the additional covariates that are included in the model.

Fig. 3 shows that there are clear income gradients in the three lifestyle indicators and SAH at most stages of the adult life course. The two notable exceptions are the lack of an income gradient in FV (Fig. 3b) during the first few years of the observed age interval and in NSMOKE (Fig. 3c) during the last few years. The income gradient in SAH (Fig. 3d) is generally stronger than the corresponding gradients in PA, FV and NSMOKE, and reaches a peak at 61 years of age, where only 48.7% of those in the first income quartile are predicted to report being in good or very good health, compared with 75.4% of those in the fourth income quartile. The fact that the income gradient in SAH is particularly strong during late midlife is even clearer in our alternative model specification, where five-year age dummies are interacted with the logarithm of income. Fig. A1 in the Appendix plots the predictions from this alternative model specification.

For the most part, the income gradients in FV, NSMOKE and SAH are qualitatively similar in that they are concave in age. That is, the gradients in these variables are stronger in midlife than during earlier and later stages of the adult life course. As discussed, this life course pattern in SAH of cumulative advantage effects in health by income until late midlife followed by age-as-leveler effects at older ages is evident in several earlier studies (Beckett, 2000; Huijts *et al.*, 2010; van Kippersluis *et al.*, 2010). What we add in this analysis is the potential role of health affecting lifestyles, such as FV and NSMOKE, in partly explaining this finding. If the income gradients in these and other important health affecting lifestyles become smaller as people age, it seems reasonable to assume that this would also hold for the income gradients in SAH and the other health indicators.

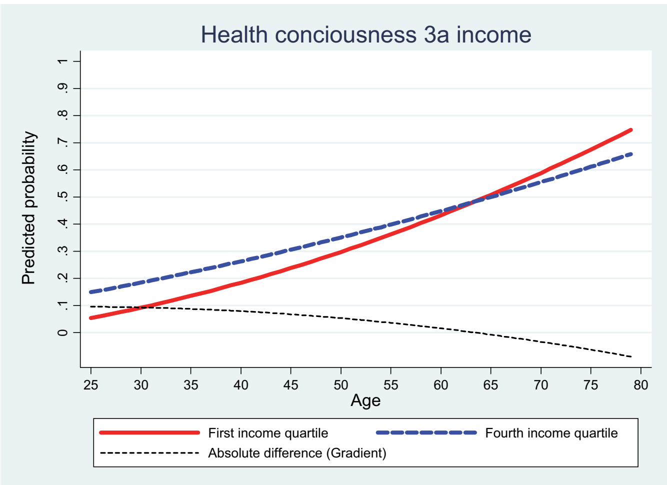


Fig. 4. Predicted age trajectories in subjective health consciousness for people in the first and fourth income quartiles. Predictions based on Model 3a and calculated at the mean values of the additional covariates that are included in the model. The dependent variable is coded one if the respondent 'totally agrees' with the statement "I always try to live healthy and keep myself in good physical condition", and zero if the respondent 'partly agrees', 'partly disagrees' or 'totally disagrees'. The underlying linear probability model is based on 21,287 observations, as 419 respondents did not respond to the question on subjective health consciousness.

Reduced income differences in lifestyles at older ages may partly reflect the role of health consciousness. We illustrate this in Fig. 4. We base the plot in this figure on Model 3a, however, the dependent variable now relates to subjective health consciousness. This variable is coded one if the respondent ‘totally agrees’ with the statement “I always try to live healthy and keep myself in good physical condition” (30.1% of the sample), and zero if the respondent ‘partly agrees’, ‘partly disagrees’ or ‘totally disagrees’ with this same statement. Not surprisingly, people become increasingly health conscious as they age. More interestingly, this process of increased health consciousness in age appears more pronounced for people in the first income quartile than in the fourth income quartile. The predicted association between income and health consciousness actually changes from positive to negative at 64 years of age and remains negative thereafter. Thus, increased health consciousness and associated lifestyle improvements in age among low income people may contribute to age-as-leveler effects in health by income, or at least to slowing down the process of cumulative advantage effects at older ages. However, this conclusion may not hold for several reasons.

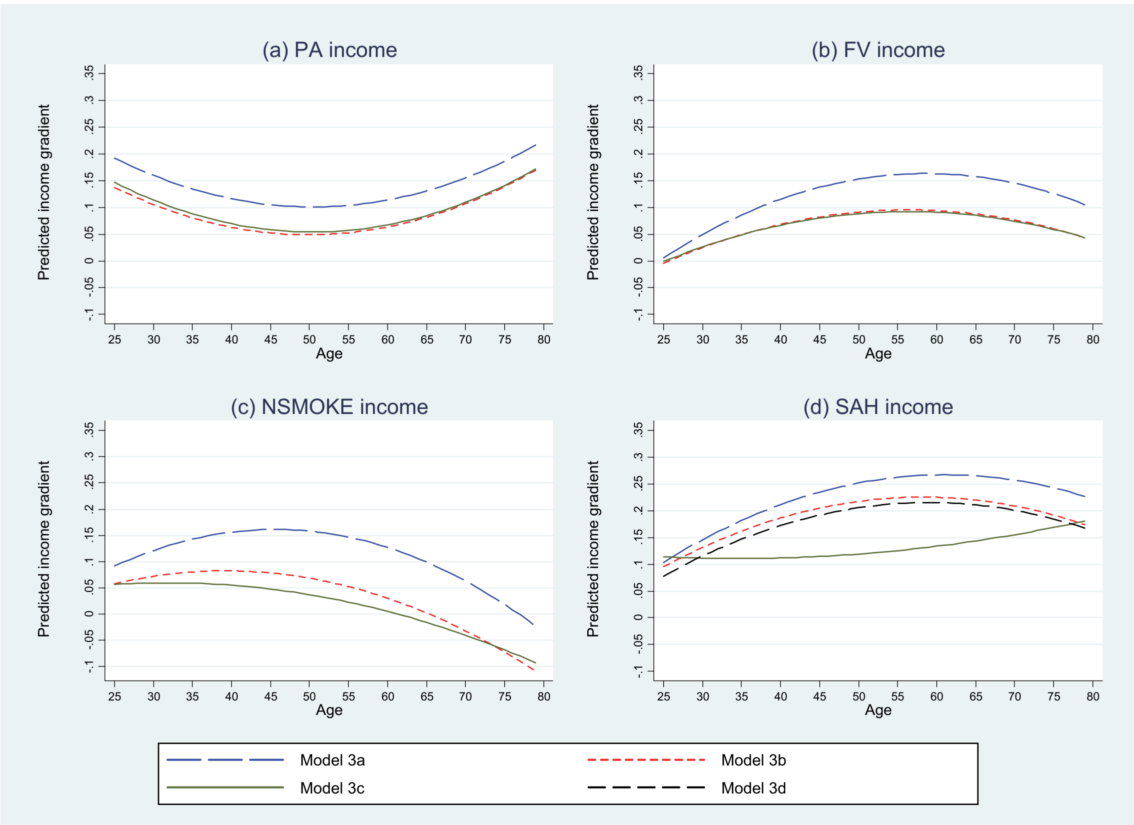


Fig. 5. Predicted age trajectories for the income gradients in lifestyles and self-assessed health resulting from Models 3a–3d in Table 2. Lines indicate absolute differences in predicted probabilities between people in the first and fourth income quartiles when controlling for different sets of variables in the linear probability models. Predictions calculated at the mean values of the additional covariates that are included in the different models.

First, Fig. 5 provides the predicted age trajectories for the income gradients in lifestyles and SAH resulting from Models 3a–3d in Table 2. As shown, the income gradients in the three lifestyle indicators are very sensitive to the choice of control variables. For example, when moving from Model 3a to Model 3b, i.e., when adding control variables for education, marital status and having children, the age-specific income gradients in PA, FV and NSMOKE are on average reduced by 40.4%, 43.0% and 60.7%, respectively. Further analysis suggests that these reductions are mainly attributable to controlling for education.

Second, as shown in Figs. 3a and 5a, the income gradient in PA is generally increasing in age at older ages. This suggests that the pattern of reduced income differences at older ages found for FV and NSMOKE does not hold for all lifestyles. However, when we estimate the PA models separately by gender, we find that the income gradient in PA is decreasing in age among males (Fig. A2a), but increasing in age among females at older ages (Fig. A3a). Thus, at least for males, it seems that also the income gradient in PA is decreasing in age at older ages.

Third, as shown in Fig. 5d, the life course pattern for the income gradient in SAH changes completely once we control for occupational status. In effect, the life course pattern changes from cumulative advantage effects in health by income until late midlife followed by age-as-leveler effects at older ages (Models 3a and 3b), to continuing cumulative advantage effects throughout the adult life course (Model 3c). Additional analysis suggests that this sensitivity of the income gradient in SAH to controlling for occupational status is almost entirely due to the effect of being reliant on social security payments during the last few years before expected retirement. On average, compared with a nonmanual worker, being on social security payments reduces the predicted probability of reporting to being in good or very good health by 39.9 percentage points. As a point of comparison, being an unskilled manual worker reduces the probability of being in good or very good health by only 4.7 percentage points.

For the 8.8% of the total sample that are on social security payments, 53.0% are in the age range 55–66 years (the official retirement age in Norway is 67 years), of which 53.4% belong to the first income quartile. Thus, as in studies from the US (Case and Deaton, 2005) and the Netherlands (van Kippersluis *et al.*, 2010), we find that in Norway, the spike in the income gradient in SAH in late midlife may largely reflect the effect of poor health on premature exit from the labor force. This in turn affects income negatively because of the shift from earning wages to being reliant on social security payments.

Finally, the age-specific income gradient in SAH is reduced by only 6.6% on average when we add our lifestyle indicators as control variables in the SAH model, i.e., when we move from Model 3b to Model 3d in Fig. 5. Because of our use of repeated cross-section data, we are unable to control for the dynamic nature of health production. That said, current lifestyles do not seem very important in moderating the current relationship between income and SAH.

4.2. Education, lifestyles and SAH over the adult life course

Table 3 presents the parameter estimates from the models focusing on education gradients in lifestyles and SAH. These model specifications are the same as in Table 2 except that the income dummies (I_2, I_3 and I_4) are replaced by education dummies (E_2, E_3 and E_4).

The results from Model 1a suggest that there are clear overall education gradients in the three lifestyle variables as well as SAH. Unlike the above findings for income, it is not clear that the education gradient in self-assessed health is steeper than the education gradients in underlying, health affecting lifestyles. In fact, the largest educational differences are found in cigarette smoking. On average, people with a university or college degree are 23.0 percentage points less likely to be daily smokers than those who have completed only lower secondary education or less.

Table 3
Linear probability models for the association between age, education, lifestyles and SAH.

Model	(1a)	(2a)	(3a)	(3b)	(1a)	(2a)	(3a)	(3b)	(3c)	(3d)
	Physical activity (PA) models				*Nonsmoking (NSMOKE) models*					
A	0.0137	0.0026	0.0562	***0.0758***	0.0040	*0.0424*	0.0128	0.0150	0.0256	
A^2	0.0018	0.0022	–0.0076	***–0.0130***	**0.0100**	**0.0092**	***0.0144***	**0.0149**	***0.0115***	
E_2	**0.0576**	0.0206	***0.0646***	0.0518	**0.0577**	**0.1519**	**0.1361**	**0.1339**	**0.1218**	
$A \cdot E_2$		*0.0162*	–0.0390	–0.0312		**–0.0308**	–0.0174	–0.0183	–0.0267	
$A^2 \cdot E_2$			*0.0114*	0.0095			–0.0021	–0.0025	0.0002	
E_3	**0.1308**	**0.1177**	**0.1722**	**0.1467**	**0.1533**	**0.2862**	**0.2551**	**0.2519**	**0.2274**	
$A \cdot E_3$		0.0021	***–0.0759***	***–0.0645***		**–0.0536**	–0.0074	–0.0139	–0.0228	
$A^2 \cdot E_3$			***0.0172***	***0.0147***			–0.0103	–0.0094	–0.0061	
E_4	**0.1581**	**0.1335**	**0.1739**	**0.1379**	**0.2303**	**0.3793**	**0.3515**	**0.3435**	**0.3132**	
$A \cdot E_4$		0.0088	–0.0374	–0.0195		**–0.0671**	–0.0255	–0.0331	–0.0415	
$A^2 \cdot E_4$			0.0091	0.0049			–0.0095	–0.0084	–0.0048	
R^2	0.0326	0.0329	0.0333	0.0402	0.0654	0.0702	0.0705	0.0825	0.0873	
	Fruits and vegetables (FV) models				*Self-assessed health (SAH) models*					
A	**0.1223**	**0.1095**	**0.1259**	**0.1300**	**–0.0759**	**–0.0817**	***–0.0838***	–0.0530	–0.0027	*–0.0633*
A^2	**–0.0161**	**–0.0157**	**–0.0189**	**–0.0189**	***0.0085***	***0.0087***	0.0088	0.0018	–0.0087	0.0019
E_2	**0.0594**	0.0183	0.0381	0.0370	**0.0708**	***0.0535***	*0.0581*	0.0456	0.0426	0.0284
$A \cdot E_2$		***0.0173***	–0.0125	–0.0152		0.0072	–0.0035	–0.0073	–0.0380	–0.0026
$A^2 \cdot E_2$			0.0066	0.0066			0.0027	0.0029	*0.0105*	0.0021
E_3	**0.1282**	**0.1017**	**0.1065**	**0.1062**	**0.1402**	**0.1379**	**0.1228**	**0.0980**	**0.0905**	*0.0606*
$A \cdot E_3$		0.0087	0.0116	–0.0008		–0.0016	0.0305	0.0162	–0.0243	0.0233
$A^2 \cdot E_3$			–0.0016	0.0004			–0.0081	–0.0059	0.0036	–0.0064
E_4	**0.1670**	**0.1370**	**0.1513**	**0.1453**	**0.1885**	**0.1698**	**0.1708**	**0.1306**	**0.1173**	**0.0851**
$A \cdot E_4$		0.0107	–0.0077	–0.0217		0.0085	0.0046	–0.0087	*–0.0519*	–0.0036
$A^2 \cdot E_4$			0.0038	0.0058			0.0011	0.0025	***0.0134***	0.0027
R^2	0.0894	0.0897	0.0899	0.1017	0.0647	0.0648	0.0651	0.0868	0.1310	0.1060

Notes: All models control for gender, survey years and birth cohorts. Models 3b–3d also control for income, marital status and having children. In addition, Model 3c controls for occupational status, while Model 3d controls for PA, FV and NSMOKE. A denotes age (centered at 30 years of age), and E_2, E_3 and E_4 denote education levels at upper secondary education, some university or college and university or college degree, respectively (the reference group is lower secondary education or less (E_1)). See Table 1 for further variable definitions. Parameters involving A and A^2 are multiplied by 10 and 10^2, respectively. Parameters in **bold**, ***bold italics*** and *italics* are statistically significant at the 99%, 95%, and 90% levels using robust standard errors, respectively. Sample weights are applied. All models are based on 21,706 observations.

There is less significant life course variation in the education gradients in lifestyles and SAH than in the corresponding income gradients. The exception is cigarette smoking, where educational differences are clearly decreasing in age (Model 2a). To study further the results in Table 3, we now turn to graphical analysis, similar to the analysis for income. To make the comparison of education and income gradients clearer, we construct Figs. 6–8 for education to be equivalent to Figs. 3–5 for income. Thus, using the results from Model 3a in Table 3, Fig. 6 shows the predicted age trajectories for the probabilities of PA, FV, NSMOKE and SAH for those who have completed lower secondary education or less and for those with a university or college degree, along with the absolute differences in predicted probabilities between these two education groups. We refer to these differences as the education gradient. Fig. 7 depicts the corresponding age trajectories in subjective health consciousness. Finally, Fig. 8 illustrates the predicted education gradients in lifestyles and SAH resulting from Models 3a–3d in Table 3, that is, from models that include different sets of control variables.

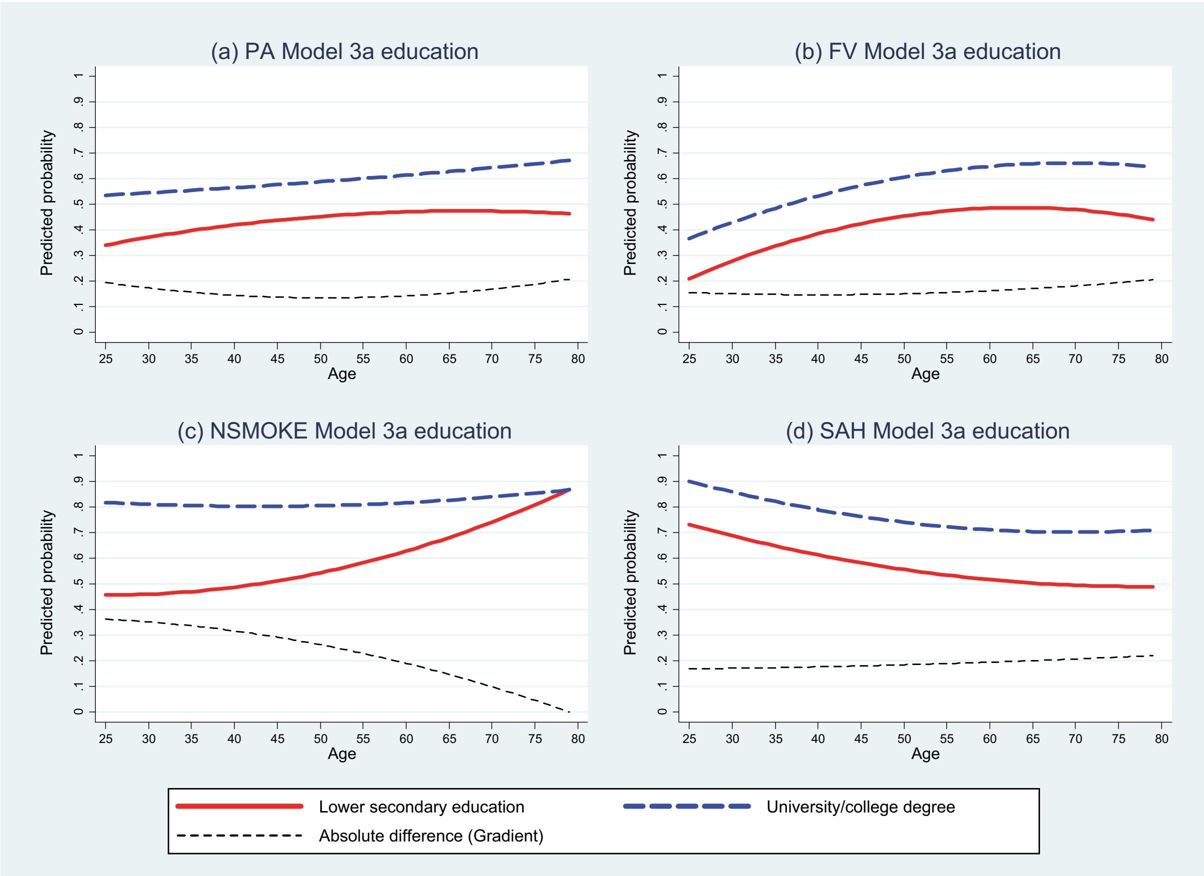


Fig. 6. Predicted age trajectories for lifestyles and self-assessed health for people in the lowest and highest education groups. Predictions based on the results of Model 3a in Table 3 and calculated at the mean values of the additional covariates that are included in the model.

Fig. 6c shows that the education gradient in NSMOKE is very steep at young ages but moves gradually towards zero at older ages. At 25 years of age, those with a university or college degree are 36.2 percentage points less likely than those that have only completed lower secondary education or less to be daily smokers. In contrast, the education gradients in PA and FV remain relatively stable throughout the adult life course. However, when we estimate the FV models separated by gender, we find that the education gradient in FV increases in age among males (Fig. A5b) but decreases in age among females at older ages (Fig. A6b). There are also very large gender differences in the predicted probabilities of eating fruits and vegetables at least two times per day. At all stages of the adult life course, the predicted probability of FV is higher among females that have completed lower secondary education or less (Fig. A6b) than among males with a university or college degree (Fig. A5b). Gender differences in the education gradient are also evident in PA, but here the pattern is opposite to that found in FV. After 55 years of age, the education gradient in PA increases in age among females (Fig. A6a) and decreases slightly in age among males (Fig. A5a).

The education gradient in SAH remains relatively stable throughout the adult life course, although it increases slightly and almost linearly in age, as shown in Fig. 6d. However, as indicated by the results in Table 3, this age variation is not statistically significant. Thus, although there are significant educational differences in SAH at all stages of the adult life course, the evidence on cumulative advantage effects in SAH by education are at most modest.

The life course patterns for the income (Fig. 4) and education (Fig. 7) gradients in subjective health consciousness are very similar, although the reduction of the gradient in age is slightly clearer in education than in income. The education gradient is also somewhat less sensitive to the addition of more control variables to the models (results not shown). These reduced educational differences in subjective health consciousness in age (Fig. 7) are reflected in 'objective' health consciousness in the case of cigarette smoking (Fig. 6c), but not in physical activity (Fig. 6a) and the consumption of fruits and vegetables (Fig. 6b). However, the subjective measures and objective indicators (lifestyles) of health consciousness are generally similar in that they are both positively associated with age.

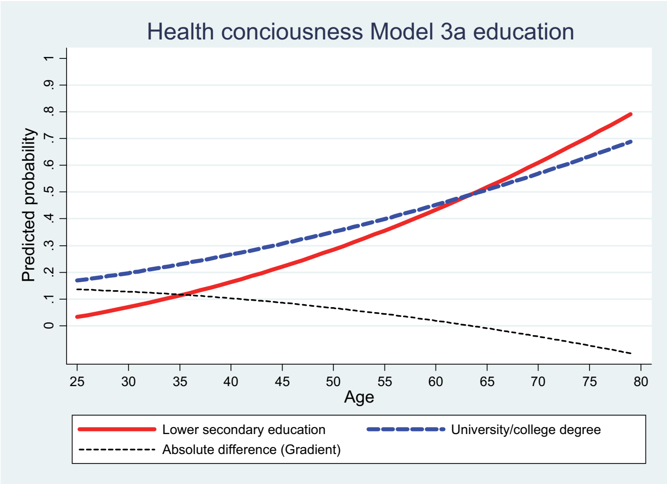


Fig. 7. Predicted age trajectories in subjective health consciousness for people in the lowest and highest education groups. Predictions based on Model 3a and calculated at the mean values of the additional covariates that are included in the model. The dependent variable is coded one if the respondent 'totally agrees' with the statement "I always try to live healthy and keep myself in good physical condition", and zero if the respondent 'partly agrees', 'partly disagrees' or 'totally disagrees'. The underlying linear probability model is based on 21,287 observations, as 419 respondents did not respond to the question on subjective health consciousness.

The education gradients in PA, FV and NSMOKE are more robust than the corresponding income gradients to adding more control variables to the models. We can see this by comparing the gradient lines for Models 3a and 3b in Figs. 5 and 8. When moving from Model 3a to Model 3b, i.e., when adding controls for income, marital status and having children, the age-specific education gradients in PA, FV and NSMOKE are on average reduced by 16.6%, 13.2% and 6.3%, respectively. As discussed, the corresponding income gradients are reduced by 40.4%, 43.0% and 60.7% when adding controls for education, marital status and having children. Thus, the positive correlation that exists between education and income appears to be important in explaining why there are income differences in PA, FV and particularly NSMOKE. Similar to the results for income, the education gradients in PA and FV are largely unaffected by controlling for occupational status, as indicated by the nearly overlapping gradient lines for Models 3b and 3c in Figs. 8a and 8b. The age-specific education gradient in NSMOKE is on average reduced by 11.0% when adding control variables for occupational status (Fig. 8c), and the gradient is more reduced at younger than older ages.

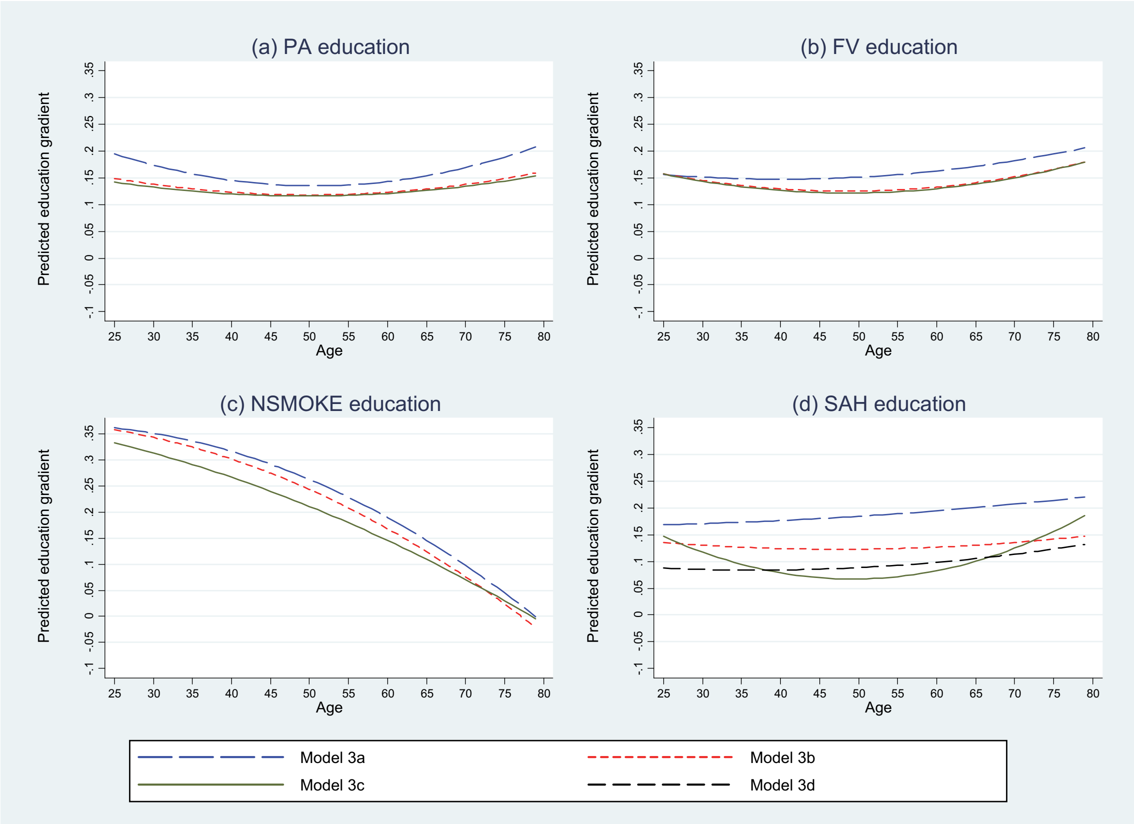


Fig. 8. Predicted age trajectories for the education gradients in lifestyles and self-assessed health resulting from Models 3a–3d in Table 3. Lines indicate the absolute differences in predicted probabilities between people in the lowest and highest education groups when controlling for different sets of variables in the linear probability models. Predictions calculated at the mean values of the additional covariates that are included in the different models.

Averaged over the adult life course, the education gradient in SAH is almost equally reduced when adding controls for occupational status (27.0%) and the lifestyle indicators (27.8%), i.e., when moving from Model 3b to Model 3c and from Model 3b to Model 3d in Fig. 8d, respectively. Thus, lifestyles seem more important in mediating the education–SAH relationship than the income–SAH relationship, as the age-specific income gradient in SAH is reduced by only 6.6% on average when adding lifestyles as control variables.

While controlling for occupational status and lifestyles almost equally affects the education gradient in SAH on average, Fig. 8d illustrates that these two factors differ in terms of their impact at different stages of the adult life course. The education gradient in SAH is moderated by lifestyles at all stages of the adult life course, and interestingly, the reduction in the gradient, i.e., the distance between the gradient lines of Models 3b and 3d in Fig. 8d, is stronger during earlier than later stages of the adult life course. The strong gradual reduction of the education gradient in NSMOKE in age (Fig. 8c) is probably important in explaining this finding. Thus, reduced educational differences in cigarette smoking at older ages could contribute to slowing down cumulative advantage effects in health by education.

Occupational status, on the other hand, is very important in explaining the education gradient in SAH during late midlife and less important during earlier and later stages of the adult life course. As for income and SAH, we find that social security status almost entirely drives this result. For the most part, we can characterize people on social security as being in poor health, in their late midlife, and clustered in the first income quartile and lowest education groups. In the last few years before expected retirement (55–66 years of age), 41.7% of those on social security have only completed lower secondary education or less, compared with 22.4% for those not on social security.

5. Discussion

The relationship between socioeconomic status and health is dynamic and evolves throughout the adult life course. Our analysis explored the role of health affecting lifestyles in explaining these dynamics. We find that in Norway, which is generally considered to be an egalitarian country (OECD, 2011), income and education are generally significantly associated with the probability of being physically active, eating fruits and vegetables and not smoking cigarettes at all stages of the adult life course.

In both low and high socioeconomic status groups, our results generally point toward increased health consciousness and associated lifestyle improvements in age as a mechanism in slowing down the natural deterioration of physical health in age. However, the predicted life course patterns for the education and income gradients in the three lifestyle indicators used in this study are too diverse to firmly conclude that this process of 'compensating behavior' at older ages is relatively stronger among lower than higher socioeconomic status groups. Thus, the role of dynamics in the relationship between socioeconomic status and lifestyles in either speeding up or slowing down cumulative advantage effects in health by income and education is not clear. At the same time, the analysis demonstrated that we should not rule out such dynamics as we find that education and income differences in lifestyles do not necessarily remain constant throughout the adult life course.

While the education gradients in physical activity and consumption of fruits and vegetables remain relatively stable throughout the adult life course, the education gradient in cigarette smoking is clearly decreasing in age after being very steep at young ages. This life course pattern in cigarette smoking appears too pronounced to be explained fully by sample selection because of high mortality rates among low-educated smokers or because of cohort effects associated with, for example, the increasing stigmatization of cigarette smokers in recent decades (Bayer, 2008). Thus, while our results generally suggest that lifestyles should contribute to cumulative advantage effects in health by education, the observed life course pattern for smoking could contribute to reducing such cumulative effects at older ages. We find some support for this mechanism in our analysis of self-assessed health.

The different patterns of results across cigarette smoking and the two other lifestyle indicators of this study may to some extent reflect systematic differences in terms of perceived health risks. That is, people with low levels of formal education quit smoking at faster rates as they age because they learn that not doing so can seriously damage their health. While eating fruits and vegetables and being physically active are also clearly associated with good health outcomes (World Health Organization, 2003), this evidence may be less accessible or perceived as less striking than the corresponding evidence on smoking.

With the exception of the income gradient in physical activity among females, the income gradients in lifestyles are generally concave in age and decreasing slightly at older ages. This could contribute to slowing down cumulative advantage effects in health by income at older ages. However, while adding our lifestyle indicators to the regression models reduces the age-specific education gradient in SAH by 27.8% on average, it reduces the corresponding income gradient by only 6.6%. To some extent, this result reflects that while the income gradients in physical activity, consumption of fruits and vegetables and particularly smoking are greatly reduced once we control for the effect of education, the reverse it not true. At least for smoking, this result seems reasonable; that is, there is no strong *a priori* reason to believe that there should be a direct causal effect running from low income to being a cigarette smoker, since the alternative (not smoking cigarettes) is less costly.

The results of this study must be considered in light of its limitations. In particular, our analysis employs repeated cross-section data, and thus we are not able to capture the dynamic nature of health production, nor are we able to capture possible feedbacks between socioeconomic status, occupational status, lifestyles and health. Thus, the results of this study are mainly of a descriptive nature, as the data generally do not allow for causal inference. Some of our key variables may also include measurement error because of incompleteness and the reliance on self-reported data, although for example SAH has been shown to be highly correlated with several objective health measures (Idler and Benyamini, 1997).

Factors such as sample selection (Kim and Durden, 2007), the increasing importance of biological factors relative to socioeconomic status in determining health at older ages (Herd, 2006), cohort effects (Lynch, 2003) and labor market characteristics (van Kippersluis *et al.*, 2010) may be important in explaining life course patterns of cumulative advantage in health by socioeconomic status until late midlife followed by age-as-leveler effects at older ages. However, our results suggest that also dynamics in the relationship between socioeconomic status, health consciousness and associated lifestyle choices may be important. Given the results and limitations of this study, there is a need for more similar research. Studies based on long panel data that track important lifestyle and health indicators as well as socioeconomic status in the same individuals over most stages of the adult life course would be particularly relevant. Studies on other lifestyle indicators, such as alcohol use and the consumption of unhealthy foods, would also be interesting, as would further analyses of the three lifestyle indicators used in this study, but possibly using alternative variable definitions (e.g., physical activity accounting for intensity level). Finally, as our results suggest that education and income differences in subjective health consciousness are gradually decreasing in age, it would be interesting to conduct similar analyses using measures of health consciousness that are more exact.

Although income differences in lifestyles potentially play some role in explaining why there are income differences in health, including how these differences evolve over the adult life course, this seems less clear than in the case of education. Given that the education gradients in physical activity, consumption of fruits and vegetables and cigarette smoking are either stable or declining over the adult life course, policies for improved lifestyle habits should mainly target young people, and particularly young people with low levels of formal education. However, targeting these groups effectively through, for example, pricing and health information policies may be difficult. That said, our results suggest that particularly among low education groups, health consciousness is increasing in age. Thus, health information policies aimed towards making people more health consciousness at earlier stages of the adult life course may be efficient. Such health information could focus on the long-term, cumulative nature of health production and thus the importance of making healthy lifestyle choices also at younger ages.

References

Baum II CL, Ruhm CJ. 2009. Age, socioeconomic status and obesity growth. *Journal of Health Economics* **28**: 635–648.

Bayer R. 2008. Stigma and the ethics of public health: not can we but should we. *Social Science & Medicine* **67**: 463–472.

Beckett M. 2000. Converging health inequalities in later life: an artifact of mortality selection? *Journal of Health and Social Behavior* **41**: 106–119.

Case A, Deaton A. 2005. Broken down by work and sex: how our health declines. In *Analyses in the economics of aging.* Wise DA (Ed.). Chicago University Press: Chicago.

Herd P. 2006. Do functional health inequalities decrease in old age? *Research on Aging* **28**: 375–392.

Huijts T, Eikemo TA, Skalická V. 2010. Income-related health inequalities in the Nordic countries: examining the role of education, occupational class and age. *Social Science & Medicine* **71**: 1964–1972.

Idler EL, Benyamini Y. 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior* **38**: 21–37.

Kim J, Durden K. 2007. Socioeconomic status and age trajectories of health. *Social Science & Medicine* **65**: 2489–2502.

van Kippersluis H, O’Donnell O, van Doorslaer E, van Ourti T. 2010. Socioeconomic differences in health over the life cycle in an egalitarian country. *Social Science & Medicine* **70**: 428–438.

Lynch SM. 2003. Cohort and life course patterns in the relationship between education and health: a hierarchical approach. *Demography* **40**: 309–331.

O’Brien RM, Hudson K, Stockard J. 2008. A mixed model estimation of age, period, and cohort effects. *Sociological Methods Research* **36**: 402–442.

OECD. 2008. *Growing unequal? Income distribution and poverty in OECD countries*. OECD Publishing: Paris.

OECD. 2011. *Society at a glance 2011 – OECD social indicators*. OECD Publishing: Paris.

Reither EN, Hauser RM, Yang Y. 2009. Do birth cohorts matter? Age–period–cohort analyses of the obesity epidemic in the United States. *Social Science and Medicine* **69**: 1439–1448.

Ross CE, Wu CL. 1996. Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior* **37**: 104–120.

Sarma S, Thind A, Chu M. 2011. Do new cohorts of family physicians work less compared to their older predecessors? The evidence from Canada. *Social Science & Medicine* **72**: 2049–2058.

Wilson A, Shuey K, Elder G. 2007. Cumulative advantage processes as mechanisms of inequality in life course health. *American Journal of Sociology* **112**: 1886–1924.

World Health Organization. 2003. Diet, nutrition and the prevention of chronic diseases. *WHO Technical Report Series No. 916*. WHO: Geneva.

APPENDIX

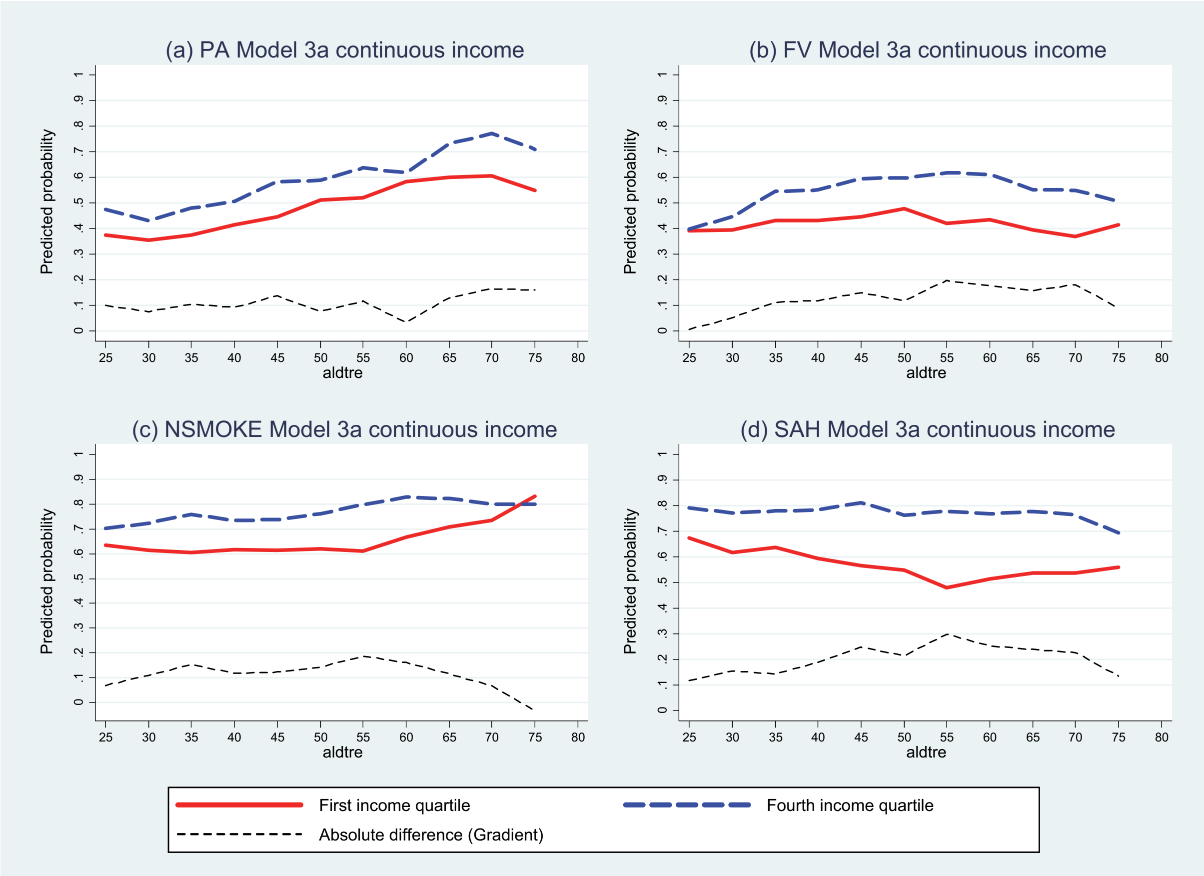


Fig. A1. Predicted age trajectories in lifestyles and self-assessed health for people in the first and fourth income quartiles based on an alternative model specification. The underlying models in this figure include interactions between five-year age dummies and the logarithm of household income. Based on the results of these models, predicted probabilities are calculated and summarized for each income quartile at each five-year age interval. The other covariates in the models are the same as in Model 3a in Table 2. Predictions are calculated at the mean values of the additional covariates.

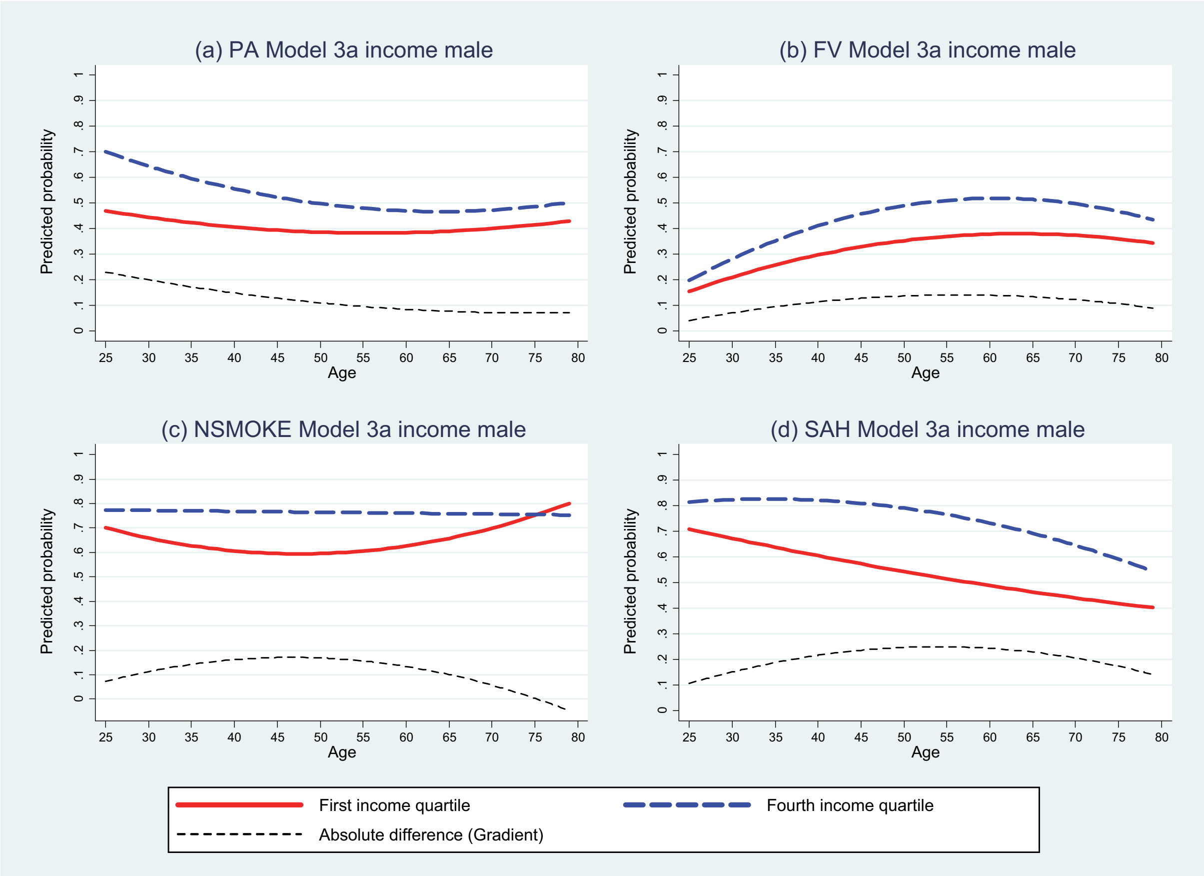


Fig. A2. Predicted age trajectories in lifestyles and self-assessed health for males in the first and fourth income quartiles. Predictions based on Model 3a applied to the male subsample and calculated at the mean values of the additional covariates that are included in the model.

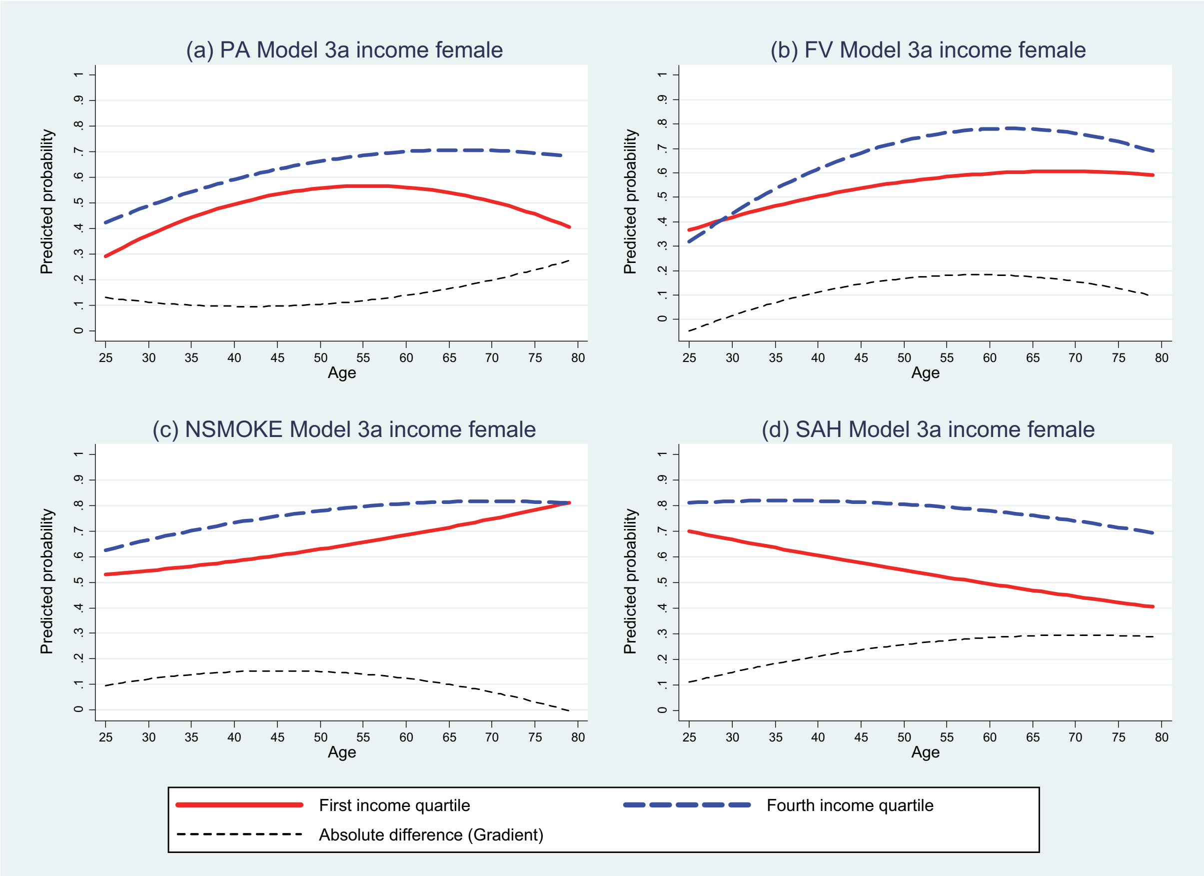


Fig. A3. Predicted age trajectories in lifestyles and self-assessed health for females in the first and fourth income quartiles. Predictions based on Model 3a applied to the female subsample and calculated at the mean values of the additional covariates that are included in the model.

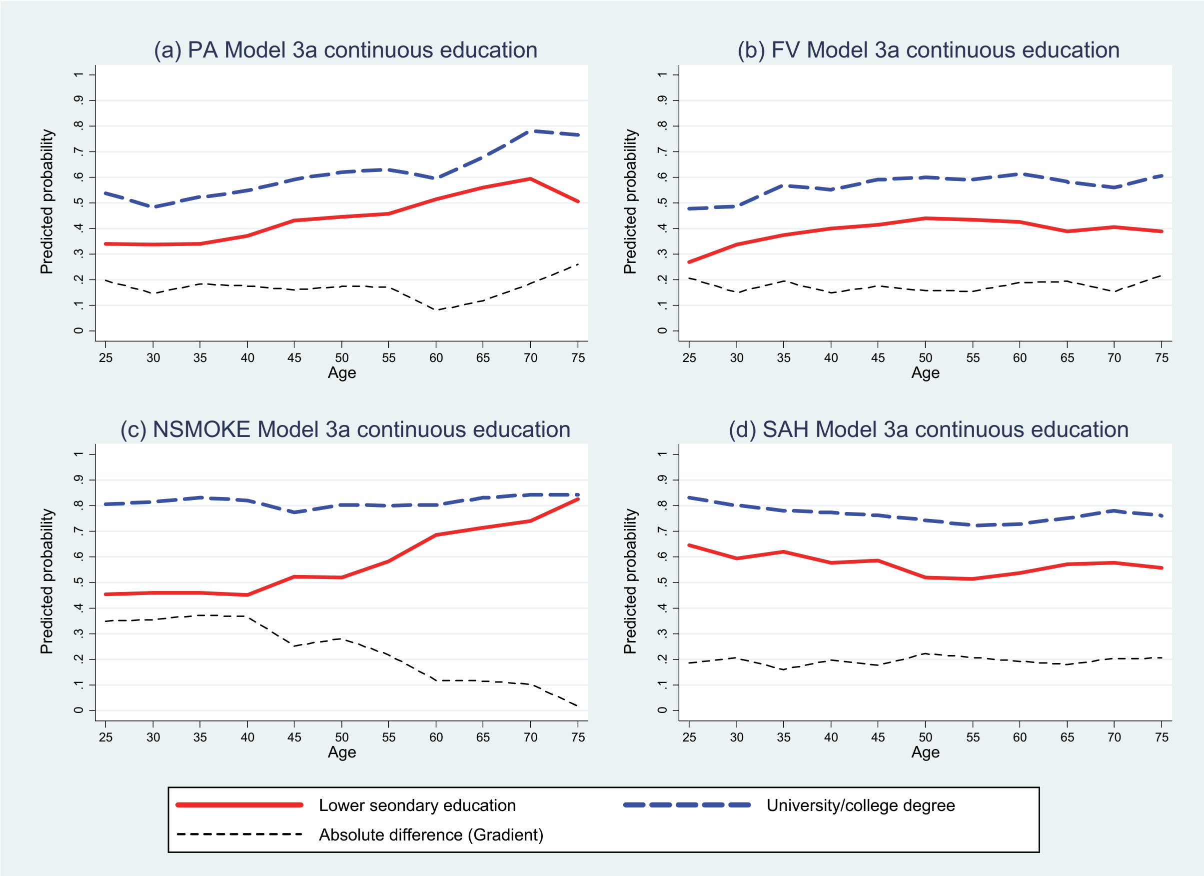


Fig. A4. Predicted age trajectories in lifestyles and self-assessed health for people in the lowest and highest education groups. The underlying models in this figure include interactions between five-year age dummies and a continuous education variable that assumes that $E_1 = 9$ years, $E_2 = 12$ years, $E_3 = 14$ years, and $E_4 = 16$ years of education. Based on the results of these models, predicted probabilities are calculated and summarized for each education group at each five-year age interval. The other covariates in the models are the same as in Model 3a in Table 3. Predictions are calculated at the mean values of the additional covariates.

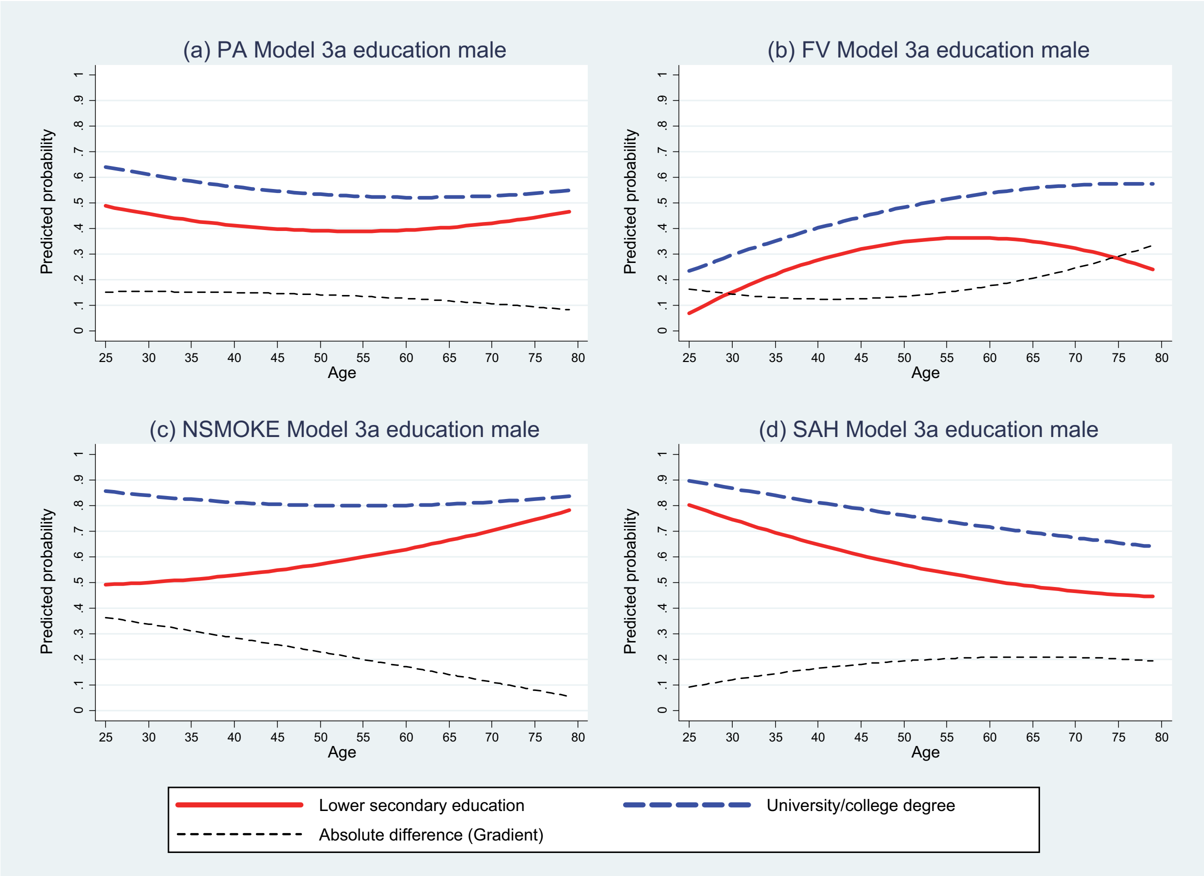


Fig. A5. Predicted age trajectories in lifestyles and self-assessed health for males in the lowest and highest education groups. Predictions based on Model 3a applied to the male subsample and calculated at the mean values of the additional covariates that are included in the model.

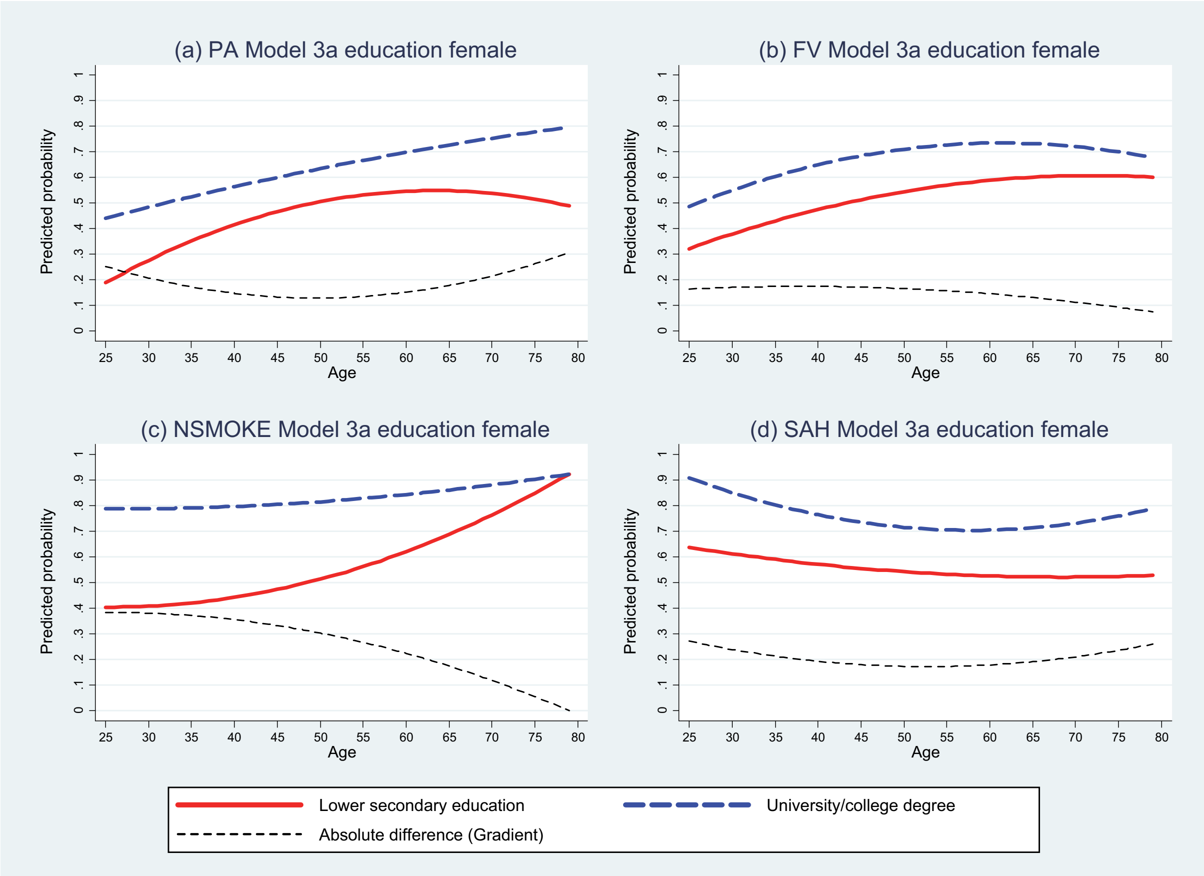


Fig. A6. Predicted age trajectories in lifestyles and self-assessed health for females in the lowest and highest education groups. Predictions based on Model 3a applied to the female subsample and calculated at the mean values of the additional covariates that are included in the model.

Paper 3

HEALTH ECONOMICS
Health Econ. **20**: 971–984 (2011)
Published online 3 October 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.1662

SOCIOECONOMIC STATUS AND LIFESTYLE CHOICES: EVIDENCE FROM LATENT CLASS ANALYSIS

ARNSTEIN ØVRUM[a,b,*]

[a]*Norwegian Agricultural Economics Research Institute, Oslo, Norway*
[b]*Department of Economics and Resource Management, Norwegian University of Life Sciences, Aas, Norway*

SUMMARY

This paper uses repeated cross-section data from Norway to estimate the demand for fruits and vegetables (FV) and physical activity (PA) with a particular focus on the role of socioeconomic status. Conventional econometric count data models produce results that are commonly found in empirical work; the effect of higher socioeconomic status on healthy behavior is positive and generally statistically significant, but the average partial effects are in some cases small and imprecisely estimated. For both behaviors, subsequent latent class models identify two subpopulations – or groups of people – with different sets of preferences; one group has low latent demands, but for these individuals, average partial effects of socioeconomic status are generally stronger than those predicted by the conventional models. The other smaller group consists of individuals who have high latent demands, but whose variability in behavior is poorly explained by socioeconomic status. Posterior analysis shows that individuals with higher socioeconomic status are more likely to belong to the healthier of these two groups. Proxies for time preferences, risk, self-control, and time constraints are also found to be important in characterizing high latent demand groups for PA and FV. Copyright © 2010 John Wiley & Sons, Ltd.

Received 10 December 2008; Revised 15 July 2010; Accepted 1 August 2010

KEY WORDS: socioeconomic status; lifestyles; physical activity; fruits and vegetables; latent class analysis

1. INTRODUCTION

People with unhealthy diets and sedentary lifestyles are at increased risk of developing chronic diseases such as type II diabetes, cardiovascular diseases and certain types of cancer (WHO, 2003; Tanasescu *et al*., 2003; Pereira *et al*., 2004; Sigal *et al*., 2004; Kay and Singh, 2006). They are also more likely to be obese, which is an intermediate risk factor for these diseases and itself a direct cause of reduced physical and mental health (Field *et al*., 2001; WHO, 2003, 2004). With obesity rates increasing in both developed and developing countries (WHO, 2004; Ogden *et al*., 2006; Hossain *et al*., 2007), and given the severe negative physiological, psychological and economic impacts of chronic disease-related morbidity and mortality – both at the individual and public level (Colditz, 1999; WHO, 2003; Yach *et al*., 2004; Yach *et al*., 2006) – it is not surprising that lifestyle research is being undertaken in many disciplines, and increasingly so.

While the key variables in statistical models for lifestyle choices and related health outcomes differ within and across disciplines, a common feature is to control for key socioeconomic variables. Although most studies confirm *a priori* expectations about positive effects of higher income and education on healthy lifestyles and good health (Johansson *et al*., 1999; Wardle and Steptoe, 2003), marginal effects

*Correspondence to: Norwegian Agricultural Economics Research Institute, P.O. Box 8024 Dep, Oslo, Norway. E-mail: arnstein.ovrum@nilf.no

are, if calculated, sometimes small or imprecisely estimated, at least for some of the included socioeconomic variables, or for some lifestyles or health outcomes in cases where more than one issue is being studied. Such mixed results for the effects of socioeconomic status on health and related behaviors are also found in carefully conducted studies such as Variyam *et al.* (2002), Chen *et al.* (2002), Contoyannis and Jones (2004), Arendt (2005), Auld and Sidhu (2005), Häkkinen *et al.* (2006), Park and Kang (2008), and Balia and Jones (2008).[1]

Several factors might explain why marginal effects of socioeconomic status on lifestyles and health are sometimes small or imprecisely estimated. First, it might be caused by erroneous or imprecise measures of relevant variables. Second, the data might just reveal true, underlying preferences; in some populations and for some lifestyle choices, the relationship between socioeconomic status and healthy behavior is not so strong and positive. And third, the effects of socioeconomic status might differ across different subpopulations within the overall population. For example, in one subpopulation, a high education level might increase health consciousness and thus physical activity (PA) levels, while in another subpopulation higher education reflects a latent preference and a relative ability for physically inactive jobs and leisure activities. Thus, unobserved preference heterogeneity clusters around a finite set of homogenous subpopulations that are unknown to the researcher *a priori*. If the population is made up of strictly distinct subpopulations, inference based on standard regression techniques may be misleading. In the above example, one might conclude that, on average, education has little impact on PA levels. While not incorrect, a more informative and precise conclusion would be that the effect is very strong for one group of people while negligible for another.

To improve and target health policies, it is important to identify such different subpopulations and their characteristics where present. In this paper, latent class models (LCMs) are being estimated to study possible subpopulation heterogeneity in the context of socioeconomic status and health-related behaviors. More specifically, the demand for PA and fruits and vegetables (FV) are being estimated using both conventional econometric count data models and their LCM counterparts. Estimating the demand for both PA and FV allows for comparing the role and the relative importance of unobserved heterogeneity across different health-related behaviors. In addition, PA and FV make good case studies, since unlike for many other lifestyle choices, there is convincing evidence for PA and FV being beneficial to health (WHO, 2003).

In LCMs – or finite mixture models – the population is viewed as a probabilistic mixture of a finite set of subpopulations, or latent classes or groups, of individuals (Cameron and Trivedi, 1998). In estimation, the log-likelihood function is specified as a weighted average of sub-distributions or component densities, of which each represent a different group or 'type' of individuals. Thus, intercept and slope heterogeneity – or utility functions – are allowed to vary across groups, but is assumed fixed within each group. The weights, which are estimated along the component densities, reflect average probabilities of belonging to the different groups.

[1]For example, studying saturated fat, cholesterol and fiber intakes among US adults by gender, Varyiam *et al.* (2002) found that across their 6 econometric models, 8 out of totally 12 education and income parameters were statistically significant at the five percent level or better, while the remaining 4 parameters were not. Chen *et al.* (2002) found income elasticities for eight healthy and unhealthy macronutrients, vitamins and minerals as well as exercising to range between −0.005–0.085, of which most were statistically insignificant. Education was not related to for example sodium and cholesterol intakes. Using Danish school reforms as instruments, Arendt (2005) was not able to find causal effects of longer education on improved self-rated health, lower BMI in the healthy range, and nonsmoking, although he finds expected correlations when education is treated as exogenous. Balia and Jones (2008) use British panel data to estimate mortality as functions of lifestyles, socioeconomic status and other controls. They find that social class is important in predicting mortality, but the effects of education on both mortality and several underlying lifestyles are less clear. Using US data and controlling for cognitive ability, Auld and Sidhu (2005) find that education affects health positively, but only at the lowest levels of schooling. Häkkinen *et al.* (2006) use Finish data and find positive effects of education on health among males, but not among females. In addition, the education effect among males was modest; five additional years of schooling was associated with only a 1% increase (0.008) on their comprehensive 15D health index score, which ranges between 0 and 1 and covers 15 different health disabilities.

LCMs have been increasingly applied in the economics literature in recent years. In health economics, LCMs have been applied on count data for health-care utilization (Deb and Trivedi, 1997, 2002; Bago d'Uva, 2005; Lourenço and Ferreira, 2005), and on ordered data for subjective well-being (Clark *et al.*, 2005). In experimental economics, LCMs have been used as an alternative or a supplement to random parameter models such as mixed logit, e.g. in transportation (Greene and Hensher, 2003), recreation demand (Boxall and Adamowicz, 2002), and recently, in health (Hole, 2008). Except for Etilé (2006), who used LCMs to study preference heterogeneity in cannabis use and heavy drinking among French teenagers, no studies have, to the best of my knowledge, applied LCMs in the context of health-related lifestyle choices. This is surprising, given the evidently complex nature of these choices; a wide variety of economic, socio-cultural, psychological and institutional triggers and constraints affect whether we choose to live healthy. *A priori*, it is natural to except that the more complex is the choice situation, the less homogenous is the population. LCMs can potentially help detect this heterogeneity and improve the way in which we estimate it. At the same time, such models might help explain why conventional models sometimes fail at finding the effects of socioeconomic status on health and related behaviors to be both statistically and substantially significant.

2. METHODS

In the latent class framework, the population is assumed to consist of C homogenous subpopulations – or classes or groups – of individuals. The probability of belonging to class j is given by π_j, with $0<\pi_j<1$, and $\sum_{j=1}^{C}\pi_j = 1$. The finite mixture density of the random variable Y is then given by

$$f_j(y_i|\boldsymbol{x_i},\boldsymbol{\theta}) = \sum_{j=1}^{C} \pi_j f_j(y_i|\boldsymbol{x_i},\boldsymbol{\theta_j}), \quad i = 1,\ldots\ldots,N \tag{1}$$

where $\boldsymbol{x}$ is a vector of characteristics for individual i and $\boldsymbol{\theta}$, which includes the π_j's, are parameters to be estimated. In this study, the random variables FV and PA are count variables. The results of count data models are potentially sensitive to the assumptions that are being made about the conditional variance of the random variable, and therefore, the class or component densities in this study will be fitted using the Poisson, negative binomial 1 (NB1) and negative binomial 2 (NB2) distributions:

$$f_j(y_i|\boldsymbol{x_i},\boldsymbol{\theta_j}) = \frac{\Gamma(y_i+\psi_{j,i})}{\Gamma(\psi_{j,i})\Gamma(y_i+1)}\left(\frac{\psi_{j,i}}{\psi_{j,i}+\lambda_{j,i}}\right)^{\psi_{j,i}}\left(\frac{\lambda_{j,i}}{\psi_{j,i}+\lambda_{j,i}}\right)^{y_i} \tag{2}$$

where $\lambda_{j,i} = \exp(\boldsymbol{x_i'\beta_j})$, $\Gamma(\cdot)$ is the gamma function and $\psi_{j,i} = (1/\alpha_j)\lambda_{j,i}^k$, where the α_j's are dispersion parameters. The Poisson model is obtained by the restriction $\alpha = 0$, while $\alpha>0$ with $k = 1$ and $k = 0$ produce the NB1 and NB2 models, respectively (Deb and Trivedi, 1997).[2]

In LCMs it is not uncommon to restrict some slope parameters to be fixed across component densities (Clark *et al.*, 2005). This could be relevant if the model has many covariates and if interest lies primarily in group heterogeneity of the constant and a few explanatory variables only. In this study, both restricted and fully unrestricted versions of the LCMs are being estimated. In the restricted models, only parameters for the income and education variables are allowed to vary across component densities, in addition to the constant terms.

The latent class probabilities of Equation (1) (π_j's) are specified using the multinomial logit model. In this study, these prior probabilities are parameterized using only constants that are restricted to sum

[2]These three alternative model specifications have conditional mean $E(y_i|\boldsymbol{x_i}) = \lambda_i$ and conditional variance $V(y_i|\boldsymbol{x_i}) = \lambda_i+\alpha\lambda_i^{2-k}$(Deb and Trivedi, 1997).

DOI: 10.1002/hec

to one over the C latent classes.[3] Posterior probabilities for component or class membership are calculated using

$$P[i \in k] = \frac{\pi_k f_k(y_i|\mathrm{x_i},\theta_\mathrm{k})}{\sum_{j=1}^{C} \pi_j f_j(y_i|\mathrm{x_i},\theta_j)} \tag{3}$$

In line with, for example, Deb and Trivedi (2002), the logarithm of $P[i \in k]$ is then regressed on a set of covariates to identify factors that are associated with the probability of respondent i belonging to class or group k. Knowledge about the correlates of class membership is important for policy relevant variables for which the estimation results reveal major differences in slope heterogeneity across classes. In addition, the predicted mean of the dependent variable is often very different across the latent groups. Consequently, a discussion of results from LCMs should focus on the calculated partial (or marginal) effects rather than on the absolute coefficient estimates. In this study, partial effects are calculated for each individual and then averaged over the sample (APE) as follows:

$$APEx_j = N^{-1} \sum_{i=1}^{N} \partial E[y_i|\boldsymbol{x_i}]/\partial \mathrm{x}_{ij} = N^{-1} \sum_{i=1}^{N} \beta_j \exp(\boldsymbol{x}_i' \boldsymbol{\beta}) \tag{4a}$$

which in the Poisson model reduces to $\beta_j \bar{y}$. For dichotomous regressors, Equation (4a) is replaced by

$$APEx_j = \mathrm{N}^{-1} \sum_{i=1}^{N} (\mathrm{E}[y_i|x_{i1},\ x_{i2}, \ldots,\ x_{ij} = 1, \ldots, x_{iK}] - \mathrm{E}[y_i|x_{i1},\ x_{i2}, \ldots,\ x_{ij} = 0, \ldots,\ x_{iK}]) \tag{4b}$$

LCMs are estimated using maximum likelihood. The log-likelihood function for FV and PA is

$$\mathrm{In}\ L = \sum_{j=1}^{N} \sum_{j=1}^{C} \mathrm{In}\ [\pi_j f_j(y_i|\boldsymbol{x_i}, \boldsymbol{\theta_j})] \tag{5}$$

where $f_i(y_i|\boldsymbol{x_i}, \boldsymbol{\theta_j})$ is given by the Poisson, NB1 and NB2 distributions in Equation (2). In this study, only results from conventional single-component models and two-component LCMs are being reported. Attempts were also made to estimate both unrestricted and restricted versions of three-component LCMs. However, these models failed to converge, which probably suggest that they are overparameterized (Deb and Trivedi, 2002).[4] The relative performance of the different models will be assessed using the Akaike and Bayesian Information Criteria (AIC and BIC). These penalized maximum likelihood criteria allow for assessing the relative performance of competing models that differ with respect to their total number of parameters (Cameron and Trivedi, 1998: Leroux, 1992).

[3]Some studies parameterize the prior probabilities using a set of covariates that that may be equal to or different from $\boldsymbol{x}$ (Boxall and Adomowicz, 2002).

[4]The LCMs in this study were programmed and estimated in Stata. To minimize the risk of having the models converge at local maxima, starting values were provided by combining parameter vectors from the single-component models and 50 random searches for improved starting values. In Stata, if requested, random searches for improved starting values are being conducted after a feasible set of initial parameter values have been identified, but before the algorithm for maximizing the log-likelihood function starts to iterate (Gould *et al.*, 2006). Furthermore, the LCMs were estimated using both the Newton–Raphson (NR) and the Broyden Fletcher Goldfarb Shanno quasi-Newton–Raphson (BFGS) algorithms, and also switching algorithms. Although it is probably more common to apply Expectation Maximization algorithms, several studies have applied NR and BFGS with success (Deb and Trivedi, 1997, 2002; Bago d'Uva, 2005).

DOI: 10.1002/hec

3. DATA AND DESCRIPTIVE ANALYSIS

This study uses individual-level data from the Norwegian Monitor Survey, a nationally representative and repeated cross-section survey, which has been conducted biannually since 1985. The survey covers a broad range of topics, including, among others, demographic and socioeconomic information, political preferences, stands on moral and ethical issues, self-perceived happiness and health, and lifestyle habits, including FV eating and PA. In the 1999 survey round, the questions on FV and PA were improved compared to prior survey rounds; more FV varieties were added to the food frequency tables, and, in addition to categories for the frequency of PA, one now started to record the duration of a typical PA. In this study, only data from the period 1999–2009 are therefore being used, and the sample is further restricted to only include people in the age range 20–69 years. After deleting observations with missing data for relevant variables, the econometric models for FV consumption and PA levels in this study include 17 712 and 18 834 observations, respectively.

The survey questions on FV and PA are categorical, i.e. they are not pure counts. For each of the 21 different FV varieties, the response alternatives to the frequency-of-eating question include 'daily'; '3–5 times per week'; '1–2 times per week'; '2–3 times per month'; 'about once a month'; '3–11 times per year'; 'rarer', 'never'. For PA, the survey includes one frequency question (eight categories ranging from 'never' to 'once or more per day') and one question on the duration of a typical workout (six categories ranging from 'less than 15 min' to 'more than 90 min'). In this study, these categorical questions have been transformed and combined to obtain an overall but 'non-pure' count estimate for FV in number of intakes per day, and for PA in number of hours per week.[5] This strategy for constructing dependent variables might not be ideal, but similar approaches have been used before with success (Bago d'Uva, 2005). Furthermore, competing strategies, such as constructing binary or ordered dependent variables, would likewise necessitate making several variable transformations, including, for example, decisions about partly arbitrary cut-off points. At the same time, these approaches are to a lesser extent than the chosen count variable approach able to combine and thus exploit the different sources of FV and PA information that are available in the data set being used in this study.

Table I shows that about 9.2% of the sample eat FV less than once a day, while 43.3% do not engage in physical activities. About 70.8% of the FV mass is concentrated in the range 1–4 intakes daily, while about 50.6% exercise between 1–4 h per week. Both distributions are skewed to the right with a few individuals having very high PA and FV demands. Obviously, neither of these variables are perfectly recorded; the FV data do not contain information about amounts consumed, only the frequency of intakes, while the PA data do not capture intensity levels or other calorie-burning activities such as housework and short walks.

Summary statistics for relevant variables are given in Table II. There are three categories of socioeconomic variables; household income; education; and economic situation for the respondent's family when he or she was between 10 and 15 years old. Per capita household income has been adjusted for inflation and is broken into five quintiles.[6] The education variable is also categorical, distinguishing

[5]More specifically, for each of the 21 different FV varieties, the number of FV intakes per day was estimated as follows; 1**'Daily'*+(4/7)* *'3–5 times per week'*+(1.5/7)**'1–2 times per week'*+((2.5/4)/7)**'2–3 times per month'*. Consequently, *'About once a month'* and lower response categories were treated as zero intakes. Finally, daily intakes from the 21 FV varieties were added and then rounded down to the nearest integer to obtain an overall count estimate for the daily number of FV intakes. A similar strategy was followed for PA; eight response categories for the frequency – *'Never' (0); 'Less than 1 time every second week' (0); '1 time every second week (0.5);' '1 time per week' (1); '2 times per week' (2); '3–4 times per week' (3.5); '5–6 times per week' (5.5); '7 times or more per week' (7)* – and six categories for the duration of a typical PA – *'Less than 15 min' (15); '15–30 min' (22.5); '31–45 min' (37.5); '46–60 min' (52.5); '1–1.5 hours' (75); 'More than 1.5 hours' (105)* – were combined to obtain an estimate of PA in minutes per week. Then, this estimate was divided by 60 and rounded down to the nearest integer to obtain an estimate of PA in number of hours per week.

[6]The household income variable is semicontinuous, with nine response alternatives representing different income intervals. Income has been set at middle point values of these intervals to make it continuous.

DOI: 10.1002/hec

Table I. Frequencies for physical activity levels and fruits and vegetables intakes

♯ Hours per week (PA), ♯ Times per day (FV)	Physical activity		Fruits and vegetables	
	Frequency	Cumulative	Frequency	Cumulative
0	8149	43.27	1631	9.21
1	3636	62.57	3231	27.45
2	2348	75.04	3761	48.68
3	1788	84.53	3208	66.80
4	1754	93.85	2333	79.97
5	0	93.85	1583	88.91
6	820	98.20	892	93.94
7	0	98.20	542	97.00
8	149	98.99	268	98.52
9	114	99.60	140	99.31
10	0	99.60	58	99.63
11	0	99.60	34	99.82
12	76	100.00	10	99.88
13			12	99.95
14			6	99.98
15–21			3	100.00
Total	18 834		17 712	

Note: The PA count variable was generated by combining eight categories for PA frequencies and six categories for the duration of a typical PA. PA has no observations on 5, 7, 10, and 11 h per week since no combinations of frequencies and durations resulted in these counts.

Table II. Summary statistics

Variable	Description	Mean	SD
Physical activity	PA, number of hours per week	1.53	1.97
Fruits and vegetables	FV, number of times per day	2.92	2.10
Age	Age of respondent	45.27	12.88
Female	If female: 1	0.56	0.50
Kids in household	If children is living in household: 1	0.48	0.50
(Living as) Married	If married or living as married: 1	0.72	0.45
Income quintile 1	If per capita household income in 1st quintile: 1	0.20	0.40
Income quintile 2	If per capita household income in 2nd quintile: 1	0.21	0.41
Income quintile 3	If per capita household income in 3rd quintile: 1	0.19	0.40
Income quintile 4	If per capita household income in 4th quintile: 1	0.19	0.39
Income quintile 5	If per capita household income in 5th quintile: 1	0.20	0.40
Secondary school	If highest completed education is secondary school: 1	0.14	0.35
High school	If highest completed education is high school: 1	0.37	0.48
Some college	If highest completed education is some college: 1	0.18	0.38
College with degree	If highest completed education is college with degree: 1	0.30	0.46
Childh. ec. poor	If family had economic worries when 10–15 years old: 1	0.21	0.41
Childh. ec. average	If family had enough money when cautious about spending when 10–15 years old: 1	0.64	0.48
Childh. ec. rich	If family well-endowed when 10–15 years old: 1	0.15	0.35
Not working	If not working part time or full time: 1	0.18	0.39
Trend	Survey year 1999 = 1, …, 2009 = 6	3.41	1.70

Note: Summary statistics for the PA and FV variables are based on 18 834 and 17 712 observations, respectively. Summary statistics for other variables are based on data from observations that were included in either the PA or FV models in Tables III–V, or both (18 991 observations), i.e. they do not include observations that were omitted from both the PA and FV models due to incomplete information (1002 observations).

between whether the respondent has completed secondary school or less, high school, some college or college with a degree.

4. RESULTS

Model selection criteria for the estimated PA and FV models are summarized in Table III. These criteria suggest that for both PA and FV, the two-component LCMs outperform the conventional single-component models. It seems that in the context of socio-demographic factors and health-related lifestyle choices, people group into distinct subpopulations. The complexity of these choices is thus associated with an extent of preference heterogeneity that may not be appropriately captured in conventional models.

The unrestricted NB1 LCMs perform better than the other models listed in Table III according to AIC and BIC. The results of these LCMs and corresponding single-component models for PA and FV are reported in Tables IV and V, respectively. In the single-component models, the effects of socioeconomic status on PA and FV demand are generally in accordance with *a priori* expectations, with all income, education, and childhood status parameters being positive, and with 15 out of these in total 18 parameters being statistically significant at the 5% level or better. Interestingly, controlling for current income levels and education, the economic status of the respondent's family when he or she was 10–15 years old is found to affect healthy behaviors in adulthood. While other studies have primarily focused on the importance of family background in predicting adult health (Fogel, 1994; Case *et al.*, 2002; Heckman, 2006), the above results suggest lifestyles to be one likely channel for creating such differences; childhood conditions partly influence the formation of persistent lifestyle habits, which in turn influence future health. It is also interesting to note that these results are found in Norway, which is generally held to be an egalitarian country, with its comprehensive, well-funded welfare state, which specifically seeks to avoid that childhood health and learning as well as later education and occupation decisions are pre-determined by family background.

The two-component LCMs for PA and FV identify qualitatively similar patterns of subpopulation heterogeneity: first, in the majority groups, labeled as the 'unhealthy groups' and representing 61.8 and 70.2% of the populations, respectively, individuals have on average low latent demands for PA and FV, at 0.81 h per week and 2.46 intakes per day. Second, individuals in the smaller, 'healthy groups' have considerably higher

Table III. Model selection criteria

Dependent variable	Model	df	ln L	AIC	BIC
Physical activity	Poisson single-component	16	−36 442.90	72 917.79	73 043.29
	LCM Poisson restricted	25	−31 732.56	63 515.12	63 711.21
	LCM Poisson unrestricted	33	−31 680.28	63 426.55	63 685.39
	NB1 single-component	17	−31 621.14	63 276.28	63 409.62
	LCM NB1 restricted	27	−31 462.56	62 979.12	63 190.89
	LCM NB1 unrestricted	35	−31 422.60[a]	62 915.19[c]	63 189.71[c]
	NB2 single-component	17	−31 731.38	63 496.77	63 630.10
	LCM NB2 restricted	27	−31 473.36	63 000.71	63 212.48
	LCM NB2 unrestricted	35	−31 432.08	62 934.16	63 208.68
Fruits and vegetables	Poisson single-component	16	−35 562.35	71 156.70	71 281.21
	LCM Poisson restricted	25	−35 144.75	70 339.51	70 534.06
	LCM Poisson unrestricted	33	−35 061.42	70 188.83	70 445.64
	NB1 single-component	17	−35 165.98	70 365.97	70 498.26
	LCM NB1 restricted	27	−35 097.82	70 249.64	70 459.75
	LCM NB1 unrestricted	35	−35 047.98[a]	70 165.96[b]	70 438.33[c]
	NB2 single-component	17	−35 237.23	70 508.45	70 640.75
	LCM NB2 restricted	27	−35 126.97	70 307.95	70 518.06
	LCM NB2 unrestricted	35	−35 050.04	70 170.07	70 442.44

[a]Preferred model based on ln L criteria.
[b]Preferred model based on Aikake information criteria (AIC); AIC $= -2 \ln L + 2K$, where K is the number of estimated parameters.
[c]Preferred model based on Bayesian information criteria (BIC); BIC $= -2 \ln L + K \ln N$, where N is the number of observations.

Table IV. Physical activity models

	Single-component NB1 model				Two-component latent class NB1 model							
					Class 1: the unhealthy group				Class 2: the healthy group			
	β	β_{SE}	APE	APE_{SE}	β	β_{SE}	APE	APE_{SE}	β	β_{SE}	APE	APE_{SE}
Age	0.000	(0.001)	0.000	(0.001)	0.004	(0.002)	0.003	(0.002)*	−0.003	(0.001)	−0.008	(0.003)**
Female	0.076	(0.018)	0.116	(0.028)***	0.308	(0.074)	0.243	(0.043)***	−0.102	(0.033)	−0.276	(0.100)***
Kids in household	−0.112	(0.024)	−0.171	(0.037)***	−0.151	(0.061)	−0.120	(0.057)**	−0.065	(0.040)	−0.175	(0.103)*
(Living as) Married	−0.092	(0.021)	−0.144	(0.034)***	0.032	(0.057)	0.026	(0.044)	−0.201	(0.031)	−0.564	(0.105)***
Income quintile 2	0.072	(0.031)	0.112	(0.050)**	0.090	(0.078)	0.075	(0.065)	0.080	(0.043)	0.221	(0.123)*
Income quintile 3	0.149	(0.032)	0.239	(0.054)***	0.206	(0.081)	0.178	(0.076)**	0.117	(0.048)	0.327	(0.135)**
Income quintile 4	0.141	(0.034)	0.226	(0.057)***	0.250	(0.093)	0.219	(0.086)**	0.087	(0.050)	0.242	(0.140)*
Income quintile 5	0.222	(0.037)	0.362	(0.065)***	0.431	(0.116)	0.390	(0.102)***	0.085	(0.053)	0.234	(0.149)
High school	0.161	(0.033)	0.253	(0.055)***	0.347	(0.095)	0.307	(0.086)***	0.038	(0.050)	0.102	(0.135)
Some college	0.318	(0.036)	0.542	(0.070)***	0.678	(0.117)	0.704	(0.138)***	0.060	(0.066)	0.163	(0.180)
College with degree	0.349	(0.034)	0.571	(0.063)***	0.844	(0.151)	0.808	(0.129)***	−0.018	(0.064)	−0.049	(0.172)
Childh. ec. average	0.049	(0.023)	0.074	(0.035)**	0.134	(0.056)	0.106	(0.044)**	−0.017	(0.035)	−0.046	(0.095)
Childh. ec. rich	0.099	(0.031)	0.157	(0.051)***	0.119	(0.075)	0.100	(0.072)	0.044	(0.045)	0.121	(0.124)
Not working	0.128	(0.026)	0.204	(0.044)***	0.043	(0.060)	0.035	(0.051)	0.231	(0.043)	0.668	(0.152)***
Trend	0.042	(0.006)	0.066	(0.009)***	0.065	(0.014)	0.053	(0.012)***	0.023	(0.009)	0.063	(0.022)***
Constant	−0.065	(0.061)			−1.647	(0.431)			1.133	(0.139)		
Dispersion (α)	1.732	(0.035)***			1.394	(0.170)***			0.605	(0.129)***		
Group probability (π)					0.618	(0.061)***			0.382	(0.061)***		
Predicted mean	1.527	(0.326)			0.808	(0.362)			2.692	(0.444)		
Log-likelihood	−31 621.14				−31 422.60							
N	18 834				18 834							

Notes: The dependent variable is physical activity in number of hours per week. White robust standard errors (β_{SE}) in parentheses. Average partial effects (APE) are calculated according to Equation (4a) and (4b), and their standard errors (APE_{SE}) are calculated using the delta method. ***Significant at the 1% level, **5% level, *10% level. Reference categories are *Income quintile 1*, *Secondary school*, and *Childh ec. poor*. Predicted means (with SD in parentheses) represent average number of PA hours per week based on posterior predictions from the estimated models.

Table V. Fruits and vegetables models

	Single-component NB1 model				Two-component latent class NB1 model							
					Class 1: the unhealthy group				Class 2: the healthy group			
	β	β_{SE}	APE	APE_{SE}	β	β_{SE}	APE	APE_{SE}	β	β_{SE}	APE	APE_{SE}
Age	0.010	(0.001)	0.030	(0.002)***	0.014	(0.001)	0.034	(0.003)***	0.005	(0.001)	0.018	(0.005)***
Female	0.386	(0.010)	1.091	(0.035)***	0.478	(0.018)	1.132	(0.057)***	0.231	(0.033)	0.903	(0.122)***
Kids in household	0.120	(0.014)	0.349	(0.040)***	0.157	(0.027)	0.388	(0.067)***	0.061	(0.037)	0.242	(0.150)
(Living as) Married	0.208	(0.013)	0.577	(0.035)***	0.318	(0.027)	0.723	(0.057)***	0.038	(0.029)	0.151	(0.113)
Income quintile 2	0.005	(0.017)	0.014	(0.049)	0.068	(0.027)	0.170	(0.070)**	−0.090	(0.039)	−0.348	(0.151)**
Income quintile 3	0.030	(0.017)	0.087	(0.051)*	0.109	(0.031)	0.278	(0.081)***	−0.090	(0.045)	−0.348	(0.169)**
Income quintile 4	0.039	(0.018)	0.115	(0.055)**	0.121	(0.033)	0.308	(0.087)***	−0.083	(0.046)	−0.321	(0.175)*
Income quintile 5	0.079	(0.021)	0.235	(0.063)***	0.164	(0.039)	0.424	(0.105)***	−0.047	(0.050)	−0.185	(0.195)
High school	0.079	(0.018)	0.232	(0.054)***	0.113	(0.029)	0.283	(0.075)***	0.018	(0.041)	0.070	(0.165)
Some college	0.162	(0.020)	0.498	(0.064)***	0.264	(0.032)	0.709	(0.097)***	0.000	(0.048)	−0.002	(0.191)
College with degree	0.228	(0.019)	0.694	(0.060)***	0.388	(0.033)	1.024	(0.099)***	−0.043	(0.054)	−0.169	(0.213)
Childh. ec. average	0.017	(0.013)	0.049	(0.037)	0.059	(0.022)	0.143	(0.052)***	−0.050	(0.030)	−0.201	(0.121)*
Childh. ec. rich	0.062	(0.018)	0.184	(0.054)***	0.071	(0.032)	0.179	(0.084)**	0.048	(0.046)	0.194	(0.190)
Not working	−0.035	(0.015)	−0.101	(0.043)**	−0.069	(0.026)	−0.167	(0.061)***	0.029	(0.036)	0.118	(0.147)
Trend	0.022	(0.003)	0.063	(0.009)***	0.018	(0.005)	0.044	(0.014)***	0.028	(0.008)	0.109	(0.031)***
Constant	−0.099	(0.035)			−0.760	(0.086)			0.974	(0.124)		
Dispersion (α)	0.322	(0.016)***			0.000	(0.000)			0.284	(0.048)***		
Group probability (π)					0.702	(0.038)***			0.298	(0.038)***		
Predicted mean	2.910	(0.742)			2.458	(0.856)			3.97	(0.185)		
log-likelihood		−35 165.98				−35 047.98						
N		17 712				17 712						

Notes: The dependent variable is fruits and vegetables consumption in number of intakes per day. White robust standard errors (β_{SE}) in parentheses. Average partial effects (APE) are calculated according to Equation (4a) and (4b), and their standard errors (APE_{SE}) are calculated using the delta method. ***Significant at the 1% level, ** 5% level, *10% level. Reference categories are *Income quintile 1*, *Secondary school*, and *Childh ec. poor*. Predicted means (with SD in parentheses) represent average number of FV intakes per day based on posterior predictions from the estimated models.

latent demands, at 2.69 h per week and 3.97 intakes per day.[7] And third, while the effects of socioeconomic status on PA and FV demand are generally small and imprecisely estimated in the smaller, healthier groups, at least for education, they are positive and statistically significant in the unhealthy groups.

In fact, the average partial effects of higher education levels are considerably higher in the unhealthy LCM groups than in corresponding single-component models. For example, in the unhealthy PA group, individuals with a college degree are predicted to exercise 48.5 min more per week than individuals with no formal education, while in the corresponding single-component model this estimate is 34.3 min per week. Similarly, in the unhealthy FV group, a college degree is associated with 7.17 more FV intakes per week, compared to 4.86 more FV intakes in the single-component model. With respect to income, average partial effects are similar across the single-component model and the two LCM groups for PA, although in the healthy group only one out of four income dummies are statistically significant at the 5% level. For FV, the average partial effects of the three highest income quintiles are between 1.8 and 3.2 times higher in the unhealthy LCM group than in the single-component model, and they are also more precisely estimated.

Note that although the average partial effects of higher socioeconomic status are generally bigger in the unhealthy LCM groups than in corresponding single-component models, they are generally not insubstantial in these latter, more conventional models either, at least when looking at the top income and education levels. For example, on average, according to the single-component models, individuals in the top income quintile exercise 21.7 min more per week and have 1.65 more FV intakes per week than individuals in the first quintile. Accumulated over time, these differences may have an impact on health. Thus, in this study, the conventional models pick up reasonably well the importance of socioeconomic status in predicting PA and FV, since income and education are important predictors of healthy behavior among individuals in the majority groups of the population, as indicated by the LCM results in Tables IV and V.

Keeping this in mind, the above results suggest that for both PA and FV – two important but different-in-nature health behaviors – there exists a majority group in the population in which latent demands for healthy behaviors are low, and in which the socioeconomic gradient is generally steeper and thus more severe than suggested by conventional econometric models for the population-averaged individual. Average partial effects of socioeconomic status in conventional models are 'attenuated' by a healthy minority group in which variation in PA and FV demand is poorly explained by income and education. This pattern of subpopulation heterogeneity identified by the LCMs may help explain why conventional econometric models for good health or healthy behaviors sometimes produce small or imprecise marginal effects of some socioeconomic status variables. The policy implication of this result is that socioeconomic status, in particular education, is indeed important – and perhaps more important than previously assumed – in predicting healthier behavior in the majority groups of the population that ideally should exercise more and eat more FV.

A number of alternative LCM specifications for PA and FV have been estimated to check for robustness of the results of the unrestricted NB1 LCMs in Tables IV and V, including the other NB1, NB2, and Poisson LCMs listed in Table III, unrestricted NB1 LCMs using continuous income and education variables, unrestricted NB1 LCMs in which PA and FV have been top coded at 10 due to suspicion of over-reporting by some respondents, and an unrestricted NB1 LCM using PA in frequencies per week rather than in number of hours per week.[8] The sizes of the unhealthy LCM groups

[7]The average number of PA hours per week and the average number of FV intakes per day in the unhealthy and healthy LCM groups represent the means of individual count predictions, which in turn are based on parameter estimates from the LCMs in Tables IV and V.

[8]Attempts were also made to estimate logit and probit LCMs using binary PA and FV variables. Results from such models could be used to assess the extent to which socioeconomic status has different effects on the probability of reaching a lower threshold value for healthy behavior than on the actual number of PA hours and FV intakes. However, these models failed to converge, which is probably due to insufficiently rich data, with a combination of binary dependent variables and only one observation point per respondent. In practice, one needs panel data in order to identify LCMs with binary dependent variables (Bago d'Uva, 2005).

DOI: 10.1002/hec

as well as the magnitude of the average partial income and education effects vary across these alternative specifications, in some cases significantly.[9] However, qualitatively, except when PA is defined in frequencies per week, all the specifications identify the same key pattern of results that was found in Tables IV and V, with one major, unhealthy group in which socioeconomic status is very important in explaining PA and FV demand, and one healthy group in which variation in demand is poorly explained by socioeconomic status. This pattern of results is also found when PA is defined in frequencies per week, but in this specification, the unhealthy group represents the minority of the population (about 30.2%), which is different from other specifications.

It should be emphasized that this study uses repeated cross-sectional data, which in general do not allow for causal inference. The results of this study must, therefore, mainly be considered as representing new and tentative associations between socioeconomic status and health related behaviors. To allow for causal inference, similar studies using panel or experimental data are needed. Finally, while the above results indicate that patterns of subpopulation heterogeneity are relatively similar for PA and FV, they may be different for other, perhaps unhealthier lifestyles, such as smoking and excessive alcohol consumption. Again, more studies are needed before conclusions can be drawn about generalization of results.

4.1 Posterior analysis – what characterize people in the healthy LCM groups?

The discussion thus far has focussed on the effects of socioeconomic status on variation in PA and FV demands, conditioned on having low and high latent demands in the first place. Table VI presents results from two ordinary least squares regressions in which the logarithm of posterior probabilities for belonging to the healthy LCM group for PA and FV have been regressed on a set of covariates. Again, the results of these regressions should not be inferred as casual relationships, but rather as an assessment of factors that are correlated with higher probabilities of belonging to the healthier, pre-estimated PA and FV groups.

The posterior regressions in Table VI include the same basic set of regessors as in Tables IV and V.[10] To possibly learn more about characteristics of the healthy PA and FV groups, a few additional dummy variables have been included to reflect time preferences, degree of risk averseness, self-control and time constraints, which may be important factors in predicting individual health and related behaviors (Grossman, 1999; Heckman, 2007). Although potentially interesting, the included variables in Table VI on these issues are crudely measured, and results must be inferred accordingly.

While the results in Tables IV and V suggested that education plays a small role in explaining variation in PA and FV demand among individuals in the smaller, healthier LCM groups, Table VI shows that there is a clear education gradient for the probabilities of belonging to these groups. With respect to income, individuals in the top quintile are more likely to belong to the healthy PA group, while effects of other income levels are less clear. In the case of FV, individuals who are not in the lowest income quintile are more likely to belong to the healthy group, and about equally so. Thus, except for the lowest quintile, there is no clear evidence of an income gradient for probabilities of belonging to the healthy FV group. Note also that this result may help explain why higher income levels and FV intakes

[9]Excluding the specification in which PA is defined in frequencies per week, the unhealthy PA groups represent between 56–65% of the population in most of the alternative LCM specifications. However, in the restricted NB1 model and when continuous income and education variables are being used, the unhealthy PA group represents about 78% of the population. The unhealthy FV groups represent between 67.8–73.5 % of the population, except in the unrestricted and restricted Poisson specifications, where the unhealthy group represents 82.1 and 85.8% of the population, respectively. The average partial effects of a college degree on PA demand range between 21.3–63.5 min per week across the different LCM specifications (mean: 43.1 min), while the same range for the different single-component models is 23.2–34.2 min (mean: 28.7 min). The range for average partial effects of a college degree on FV demand across the different LCM specifications is 5.73–7.62 intakes per week (mean: 6.79 intakes), while across different single-component models this range is 4.58–4.85 intakes (mean: 4.72 intakes).

[10]The linear age variable in Tables IV and V has been replaced by age group dummies in Table VI to control for possible nonlinearities in age.

DOI: 10.1002/hec

Table VI. Posterior regressions for probabilities of belonging to the healthy LCM groups

	Physical activity		Fruits and vegetables	
	β	SE	β	SE
Age 30–39	−0.047	(0.025)*	0.047	(0.018)**
Age 40–49	0.043	(0.025)*	0.072	(0.018)***
Age 50_59	0.012	(0.025)	0.128	(0.018)***
Age 60_69	0.084	(0.029)****	0.176	(0.020)***
Female	0.136	(0.014)***	0.107	(0.009)***
Kids in household	−0.026	(0.019)	0.028	(0.013)**
(Living as) Married	0.133	(0.017)***	0.132	(0.012)***
Income quintile 2	−0.035	(0.022)	0.108	(0.017)***
Income quintile 3	−0.008	(0.022)	0.114	(0.016)***
Income quintile 4	0.006	(0.024)	0.109	(0.017)****
Income quintile 5	0.064	(0.026)**	0.099	(0.020)***
High school	0.060	(0.024)**	0.048	(0.018)***
Some college	0.152	(0.027)***	0.119	(0.018)***
College with degree	0.237	(0.024)***	0.185	(0.017)***
Childh. ec. average	0.048	(0.017)***	0.069	(0.012)***
Childh. ec. rich	0.032	(0.023)	0.009	(0.017)
Not working	−0.176	(0.022)***	−0.061	(0.016)**
Trend	0.006	(0.004)	−0.007	(0.003)**
Time preferences, risk, self-control, time use				
Like to pay in installments[a]	−0.085	(0.018)***	−0.016	(0.013)
Household has life insurance	0.033	(0.013)**	0.036	(0.009)***
Willing to take big risks to achieve life goals[b]	0.044	(0.013)***	0.065	(0.009)***
Feel self-control over life outcomes[c]	0.088	(0.019)***	0.052	(0.014)***
Work overtime at least once per week	−0.091	(0.015)***	0.044	(0.010)***
Watched TV three hours or more yesterday	−0.097	(0.017)***	−0.074	(0.012)***
Constant	−1.662	(0.042)***	−1.913	(0.032)***
R^2	0.045		0.066	
N	17 760		16 774	

Notes: The dependent variables are the log of posterior probabilities for belonging to the healthy LCM groups, which are obtained from results of the two-component PA and FV LCMs in Tables IV and V, using Equation (3). There are fewer observations in these posterior regressions than in the two-component LCMs due to the inclusion of additional regressors and therefore more observations with missing data. White robust standard errors (SE) in parantheses. ***Significant at the 1% level, **5 level, *10% level. Reference categories are *Age 20_29*, *Income quintile 1*, *Secondary school*, and *Childh ec. poor*.
[a]Respondent 'partly agrees' or 'totally agrees' in that he/she likes to purchase in instalments.
[b]Respondent 'partly agrees' or 'totally agrees' in that he/she is willing to take big risks to achieve life goals.
[c]Respondent 'partly disagrees' or 'totally disagrees' in the statement. 'It is of little use to plan for the future, since what happens in life is mostly a matter of being lucky or unlucky anyway'.

are negatively related in the healthy LCM group in Table V; Table VI shows that individuals in the lowest income quintile are not likely to belong to the healthy FV group, but the few who do belong here most likely eat a lot of FV.

The probabilities of belonging to the healthy PA and FV groups are generally correlated with the included proxies for time preferences, risk, self-control, and time constraints in the expected directions. For example, households that have signed up for life insurance, which may reflect both time preferences for health and risk averseness, are more likely to belong to the healthier PA and FV groups. In addition, more likely to belong to these groups are individuals who disagree with the statement *It is of little use to plan for the future, since what happens in life is mostly a matter of being lucky or unlucky anyway*. Thus, not surprisingly, individuals who lack a feeling of self-control over life outcomes are not likely to belong to these groups. To convince these individuals that the risk of bad health outcomes such as chronic disease can be reduced through healthy lifestyles represents a difficult but important policy challenge. Working overtime at least once per week is the only parameter among this group of variables for which signs are different in the PA and FV models, which reflects that PA is a time consuming activity while FV is not, and perhaps, that individuals working overtime compensate for lack of exercise through eating more FV.

5. SUMMARY AND CONCLUSIONS

This paper introduces latent class analysis to the context of socioeconomic status and lifestyle choices in adults using PA and FV as examples. In conventional mean-effects-type econometric models, socioeconomic status and healthy behavior are typically found to be positively related, but marginal effects are sometimes small or imprecisely estimated. The results of the LCMs in this study offers one possible explanation for such findings; for both PA and FV, the LCMs identify one group of individuals, which represents 38.2 and 29.8% of the populations, respectively, that have high latent demands, but whose variability in behavior is poorly explained by socioeconomic status. The other two, bigger groups have low latent demands, but for these individuals, average partial effects of socioeconomic status are generally stronger than predicted by conventional, mean-effects-type econometric models. Thus, for individuals in these important target groups for improved health, the socioeconomic gradient in important lifestyles may be steeper and thus more severe than previously assumed. Posterior analysis shows that individuals with higher socioeconomic status are more likely to belong to the healthier latent groups. Belonging to these groups is also found to be related to proxies for time preferences, risk averseness, self-control, and time constraints.

Although the key results of this study indicate that LCMs may represent a new and useful tool in accommodating preference heterogeneity associated with highly complex lifestyle choices, more research is needed. Several limitations of this study have been pointed out, among which its use of repeated cross-section data is most important. To allow for causal inference and to make results more operational and relevant for policy, preferences should be elicited using panel data or experimental methods, and more background data on each respondent should be collected. This would allow for a more disaggregated analysis, with potentially more than two latent groups being identified, and with a richer, more informative characterization of people who belong to different groups. If LCMs can help identify specific triggers and constraints for specific groups of people for important health affecting lifestyles, policies for improved health could potentially become more targeted and thus more efficient.

ACKNOWLEDGEMENTS

This research was supported by the Research Council of Norway, grant numbers 182289 and 184809. I thank Kyrre Rickertsen, Frode Alfnes, Joseph Herriges and two anonymous referees for their helpful comments and suggestions.

REFERENCES

Arendt JN. 2005. Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review* **24**: 149–160.

Auld MC, Sidhu N. 2005. Schooling, cognitive ability and health. *Health Economics* **14**: 1019–1034.

Bago d'Uva T. 2005. Latent class models for use of primary care: evidence from a British panel. *Health Economics* **14**: 873–892.

Balia S, Jones AM. 2008. Mortality, lifestyle and socio-economic status. *Journal of Health Economics* **27**: 1–26.

Boxall P, Adamowicz W. 2002. Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and Resource Economics* **23**: 421–446.

Cameron AC, Trivedi PK. 1998. *Regression Analysis of Count Data*. Cambridge University Press: New York.

Case A, Lubotsky D, Paxson C. 2002. Economic status and health in childhood: the origins of the gradient. *American Economic Review* **92**(5): 1308–1334.

Chen S, Shogren JF, Orazem PF, Crocker TD. 2002. Prices and health: identifying the effects of nutrition, exercise, and medication choices on blood pressure. *American Journal of Agricultural Economics* **84**(4): 990–1002.

Clark A, Etilé F, Postel-Vinay F, Senik C, van der Straeten K. 2005. Heterogeneity in reported well-being: evidence from twelve European countries. *The Economic Journal* **115**: C118–C132.

DOI: 10.1002/hec

Colditz G. 1999. Economic costs of obesity and inactivity. *Medicine and Science in Sport and Exercise* **31**(Suppl. 11): S663–S667.

Contoyannis P, Jones AM. 2004. Socio-economic status, health and lifestyle. *Journal of Health Economics* **23**: 965–995.

Deb P, Trivedi PK. 1997. Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics* **12**: 313–336.

Deb P, Trivedi PK. 2002. The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics* **21**: 601–625.

Etilé F. 2006. Who does the hat fit? Teenager heterogeneity and the effectiveness of information policies in preventing cannabis use and heavy drinking. *Health Economics* **15**: 697–718.

Field AE, Coakley EH, Must A, Spadano JL, Laird N *et al*. 2001. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Archives of Internal Medicine* **161**: 1581–1586.

Fogel R. 1994. Economic growth, population theory, and physiology: the bearing of long-term processes on the making of economic policy. *American Economic Review* **84**(3): 369–395.

Gould W, Pitblado J, Sribney W. 2006. *Maximum Likelihood Estimation with Stata* (3rd edn). Stata Press: College Station, TX.

Greene WH, Hensher D. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B* **37**: 681–698.

Grossman M. 1999. The human capital model of the demand for health. *NBER Working Paper 7078*.

Häkkinen U, Järvelin MR, Rosenqvist G, Laitinen J. 2006. Health, schooling and lifestyle among young adults in Finland. *Health Economics* **15**: 1201–1216.

Heckman JJ. 2006. Skill formation and the economics of investing in disadvantaged children. *Science* **312**: 1900–1902.

Heckman JJ. 2007. The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences* **104**: 13250–13255.

Hole AR. 2008. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics* **27**: 1078–1094.

Hossain P, Kawar B, El Nahas M. 2007. Obesity and diabetes in the developing world: a growing challenge. *New England Journal of Medicine* **356**(9): 213–215.

Johansson L, Thelle DS, Solvoll K, Bjørneboe GEA, Drevon CA. 1999. Healthy dietary habits in relation to social determinants and lifestyle factors. *British Journal of Nutrition* **81**: 211–220.

Kay SJ, Fiatarone Singh MA. 2006. The influence of physical activity on abdominal fat: a systematic review of the literature. *Obesity Reviews* **7**: 183–200.

Leroux BG. 1992. Consistent estimation of a mixing distribution. *Annals of Statistics* **20**(3): 1350–1360.

Lourenço OD, Ferreira PL. 2005. Utilization of public health centres in Portugal: effect of time costs and other determinants. Finite mixture models applied to truncated samples. *Health Economics* **14**: 939–953.

Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. 2006. Prevalence of overweight and obesity in the United States, 1999–2004. *Journal of the American Medical Association* **295**: 1549–1555.

Park C, Kang C. 2008. Does education induce healthy lifestyle? *Journal of Health Economics* **27**: 1516–1531.

Pereira MA, O'Reilly E, Augustsson K, Fraser GE, Goldbourt U *et al*. 2004. Dietary fiber and risk of coronary heart disease: a pooled analysis of cohort studies. *Archives of Internal Medicine* **164**: 370–376.

Sigal RJ, Kenny GP, Wasserman DH, Castaneda-Sceppa C. 2004. Physical activity/exercise and type 2 diabetes. *Diabetes Care* **27**: 2518–2539.

Tanasescu M, Leitzmann MF, Rimm EB, Hu FB. 2003. Physical activity in relation to cardiovascular disease and total mortality among men with type 2 diabetes. *Circulation* **107**: 2435–2439.

Variyam JN, Blaylock J, Smallwood D. 2002. Characterizing the distribution of macronutrient intake among US adults: a quantile regression approach. *American Journal of Agricultural Economics* **84**: 454–466.

Wardle J, Steptoe A. 2003. Socioeconomic differences in attitudes and beliefs about healthy lifestyles. *Journal of Epidemiology and Community Health* **57**: 440–443.

World Health Organization. 2003. Diet, nutrition and the prevention of chronic diseases. *WHO Technical Report Series No. 916*, Geneva, Switzerland.

World Health Organization. 2004. Obesity: preventing and managing the global epidemic. *WHO Technical Report Series No. 894*, Geneva, Switzerland.

Yach D, Hawkes C, Gould CL, Hofman KJ. 2004. The global burden of chronic diseases: overcoming impediments to prevention and control. *Journal of the American Medical Association* **291**(21): 2616–2622.

Yach D, Stuckler D, Brownell KD. 2006. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nature Medicine* **12**: 62–66.

DOI: 10.1002/hec

Paper 4

Health information and diet choices: results from a cheese experiment

Arnstein Øvrum[a,b*], Frode Alfnes[a], Valérie Lengard Almli[c] and Kyrre Rickertsen[a,b]

[a]Department of Economics and Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway
[b]Norwegian Agricultural Economics Research Institute, P.O. Box 8024 Dep, N-0030 Oslo, Norway
[c]Nofima Mat AS, Osloveien 1, N-1430 Ås, Norway

Abstract

Health information policies may help improve knowledge, raise awareness, reduce confusion and thereby make healthier food options more attractive and visible. The distributional effects of such policies across socio-demographic groups are difficult to measure and thus not well-known. This paper reports results from a stated preference experiment on everyday use semi-hard cheese from Norway. Half of the participants were exposed to diet-related health information before performing either a choice or a ranking task. The effects of health information on marginal willingness to pay for low-saturated-fat, low-fat and organic cheese are analyzed using rank-ordered mixed logit models. Non-college, medium-high income, age 50–70 and female participants are more clearly affected by health information than college, low income, age 30–49 and male participants. Subjective statements on diet-health knowledge and awareness are used to discuss these findings. Our results suggest that provision of health information may help reduce educational differences in diet-health knowledge and thus dietary behavior. Low income participants seem to be constrained by high food prices, but not by lack of knowledge or awareness. Finally, reaching out to young people and in particular males through health information policies seems difficult.

JEL classification: D12; D80; I12; I14; I18; Q18

Keywords: cheese; choice experiment; diet choices; health information; socioeconomic status

* Corresponding author. E-mail: arnstein.ovrum@nilf.no. This paper was submitted to a peer-reviewed journal in November 2011.

1. Introduction

Increasing prevalence of obesity and related chronic diseases such as type II diabetes represent key health challenges in most developed countries, and increasingly also in developing countries (Hossain *et al.*, 2007). Excessive energy intakes and poor nutrition have contributed to these trends (World Health Organization, 2003). Our daily decisions about eating healthy or unhealthy foods are influenced by a highly complex mix of factors, including structural factors such as prices and time, budget and knowledge constraints (Blaylock *et al.*, 1999); psychological factors such as self-control, motivation and time preferences for health (Loewenstein and Prelec, 1992); social factors such as the communication of group membership through dietary habits (Etilé, 2007); and neurobiological factors such as taste preferences and addictions (Camerer *et al.*, 2005). Many of these factors, and thus also dietary behavior and health, are in turn closely related to socioeconomic status indicators such as education and income (CSDH, 2008). In many countries, including Norway, reducing socioeconomic inequalities in health is identified as a key goal for health policy (Norwegian Ministry of Public Health and Care Services, 2006).

Nutrition policies can potentially target some of the above triggers and constraints for making healthful diet choices. These include health information policies such as nutritional labeling schemes, school education, official dietary guidelines and media campaigns (Nayga, 2008). The distributional effects of such information policies are difficult to predict. For example, how will different education groups respond to a public information campaign on the importance of following a healthy diet? Due to different *a priori* levels of diet-health knowledge, it seems reasonable to expect that the marginal effect of health information on preferences for healthy foods should be larger in lower than higher education groups. On the other hand, lower and higher education groups may be systematically different in their ability to process and adapt to health information (Grossman, 2000), as well as in their general interest for health information. Thus, the effects of health information may also be positively associated with years of schooling.

This paper uses data from a stated preference (SP) experiment on everyday use semi-hard cheese from Norway to examine how diet choices are affected by exposure to diet-related health information. More specifically, it focuses on to what extent such health information effects vary by socio-demographic characteristics. Half of the participants were exposed to objective health information related to cheese consumption prior to performing either a choice or a ranking task. Thus, the access to and use of health information is exogenously determined, which is different from most non-experimental settings. Using mixed logit models, we examine whether exposure to health information affects the participants' marginal willingness to pay (MWTP) for low-saturated-fat, low-fat and organic cheese, and to what extent these information effects vary by age, gender, education and income. The results from the SP experiment are then discussed in light of the participants' responses to subjective statements on diet-health knowledge and awareness.

While there is an ongoing debate about the overall health properties of dairy foods and dairy fats, including cheese, there is less disagreement in that people could benefit from choosing low-saturated-fat and low-fat variants when consuming dairy foods (World Health Organization, 2003; Norwegian Directorate of Health, 2011). Dairy products represent the leading source of saturated fats in Norwegians' diets, and while the annual per capita consumption of fluid milk decreased by about 39% in the period 1989–2008, the consumption of cheese increased by about 25% during the same period, from 13.3 kg to 16.6 kg (Norwegian Directorate of Health, 2010). In addition to constituting an important component in many consumers' diet, everyday use semi-hard cheese may represent a food for which choices among healthy and less healthy varieties are being made unconsciously or with limited knowledge about how these and similar everyday dietary choices may affect our health in the long run. Thus, as a case study, cheese should be well-suited for studying the relationship between health information and diet choices.

2. Background

Health information policies may improve knowledge, raise awareness, reduce confusion and thereby make healthier food options more attractive and visible. Public media campaigns on health behaviors have predominantly focused on tobacco use. Although these campaigns have generally been successful (Wakefield *et al.*, 2010), they have frequently been most efficient in targeting higher socioeconomic status groups (Niederpeppe *et al.*, 2008). Public media campaigns on nutrition have been less common and have usually been combined with other types of nutrition policies, and it has therefore been difficult to isolate campaign effects themselves (Wakefield *et al.*, 2010).

It is generally difficult to compare how different socio-demographic groups are affected by exposure to diet-related health information using observational data. Typically, there is lack of cross-section variation in the supply of information, and variability over time may be confounded with secular trends in dietary habits (Ippolito and Mathios, 1994). Studies focusing on the role of socio-demographic characteristics have therefore often relied on demand-driven health information indicators such as diet-health knowledge and awareness, and the use of nutrition labels. However, results from studies that account for the likely endogeneity of such variables are mixed (Park and Davis, 2001; Variyam, 2008). And while we might find, for example, that education is positively correlated with diet-health knowledge and awareness, and that education is in turn positively correlated with making healthful diet choices (Wardle *et al.*, 2000), this has few direct policy implications. Instead, policy makers are interested in knowing whether public provision of diet-related health information can help reduce educational differences in knowledge and awareness, and thereby reduce educational differences in dietary behavior.

Some of the above limitations may be overcome by utilizing properties of controlled experiments (Roosen and Marette, 2011). Experimental studies on health information and food choices have predominantly focused on issues related to food safety and debated food technologies such as hormone treatments and genetic modifications (Alfnes and Rickertsen, 2003; Huffman *et al.*, 2007). Relatively few experimental studies have focused on health information in relation to obesity and diet-related chronic diseases (Nayga, 2008). However, some studies have examined the effects of placing diet-related health information on food items or restaurant menus, such as health claims (Gracia *et al.*, 2009), calorie recommendations (Wisdom *et al.*, 2010) and criteria-based nutrition labels such as the traffic light system in the UK (Balcombe *et al.*, 2010). A few studies have taken a different approach in that they have provided scripts of objective diet-related health information, either randomly to some of the participants at the beginning of the experiment (Lusk *et al.*, 2008), as in our study, or by providing more information at increments at different stages of the experiment (Roosen *et al.*, 2009). Lusk *et al.* (2008) informed some participants about the potential health benefits of eating pasture-grazed meat. This information had a significant effect on the consumers' MWTP for pasture-grazed steak, but not for pasture-grazed ground beef. Studying preferences for sardines and tuna, Roosen *et al.* (2009) provided information on both the health risks (methylmercury) and health benefits (omega-3) associated with these two fish species, and their findings suggest that people are more responsive to messages of health risks than health benefits.

While most of the above studies found that the consumers' food choices were significantly influenced by health information, relatively little is known about to what extent these effects vary systematically by socio-demographic characteristics. In Roosen *et al.*

(2009), the effects of health information were positively associated with education and negatively associated with income. The health claims effects in Gracia *et al.* (2009) did not vary significantly by socio-demographic factors, while in Balcombe *et al.* (2010), female and higher educated respondents were more clearly affected by nutritional traffic light labels than male and lower educated respondents.

3. The stated preference experiment

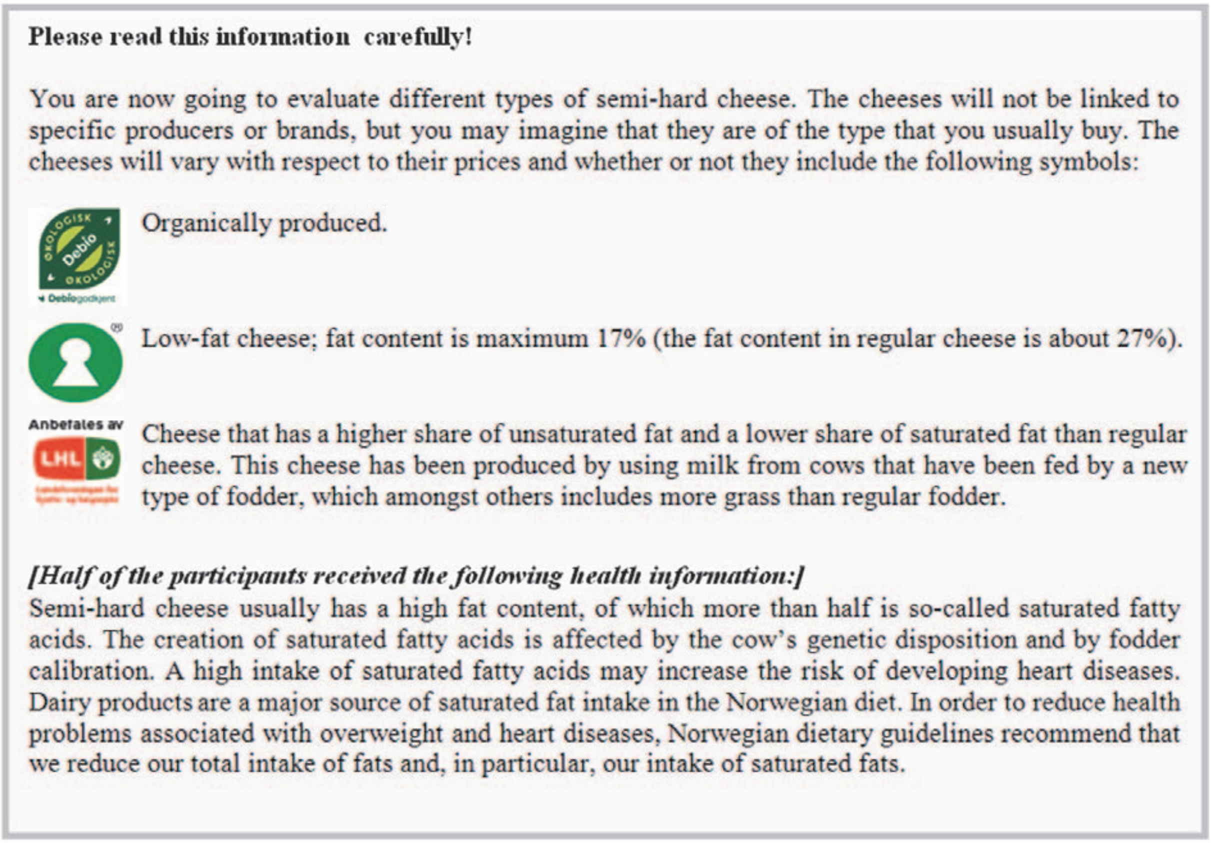
This study uses data from a SP experiment on cheese from Norway. A professional survey company (Synovate Norway) was engaged to collect the data during spring 2009. To participate in the Internet survey, the respondents had to eat cheese and buy groceries on a regular basis. The survey was completed by 426 adults in the age range 30–70 years, but only 408 observations are used here due to missing income information on eighteen respondents.

Table 1
Socio-demographic characteristics. Variable descriptions and means.

Variable	Description	Mean
College	If having attended college or university: 1; otherwise: 0	0.686
Med-high income	If not in lowest one third of household income: 1; otherwise: 0	0.684
Age 50–70	If age is 50–70 years: 1; if age is 30–49 years: 0	0.566
Female	If female: 1; if male: 0	0.559

Notes: Sample means are based on 408 observations. The mean of Med-high income is higher than 0.667 since our semi-continuous income variable had several identical values around the cut-off point at the 33.33rd percentile.

Descriptive statistics about the sample are summarized in Table 1. The variables for age, education and income are dichotomized to facilitate later empirical analyses. A high education level is defined as having attended either college or university. We distinguish between a low income group and a medium-high income group, where low income is defined as belonging to the lowest one third of a semi-continuous household income variable. The original survey question on household income included eleven response alternatives, each representing a specific income interval. To obtain our semi-continuous household income variable, we set household income to the mid-point values of each income interval, and then adjusted for household size by dividing the resulting income measure by the square root of household size (OECD, 2008). Our sample is somewhat overrepresented with college-educated, high income, older age and female respondents.


Please read this information carefully!

You are now going to evaluate different types of semi-hard cheese. The cheeses will not be linked to specific producers or brands, but you may imagine that they are of the type that you usually buy. The cheeses will vary with respect to their prices and whether or not they include the following symbols:

Organically produced.

Low-fat cheese; fat content is maximum 17% (the fat content in regular cheese is about 27%).

Cheese that has a higher share of unsaturated fat and a lower share of saturated fat than regular cheese. This cheese has been produced by using milk from cows that have been fed by a new type of fodder, which amongst others includes more grass than regular fodder.

[Half of the participants received the following health information:]
Semi-hard cheese usually has a high fat content, of which more than half is so-called saturated fatty acids. The creation of saturated fatty acids is affected by the cow's genetic disposition and by fodder calibration. A high intake of saturated fatty acids may increase the risk of developing heart diseases. Dairy products are a major source of saturated fat intake in the Norwegian diet. In order to reduce health problems associated with overweight and heart diseases, Norwegian dietary guidelines recommend that we reduce our total intake of fats and, in particular, our intake of saturated fats.

Fig. 1. Introduction screen presented to the participants at the beginning of the SP experiment (translated from Norwegian).

The generic everyday use semi-hard cheese of the SP experiment had four attributes, and each attribute had two levels: (i) a price of 42 vs. 58 Norwegian kroner (NOK) per five hundred gram cheese,[1] (ii) regular-saturated-fat vs. low-saturated-fat cheese, (iii) regular-fat vs. low-fat cheese and (iv) conventional vs. organic cheese. More information about the different cheese attributes is provided in Fig. 1, which shows the initial information screen that was presented to the participants at the beginning of the experiment. Pictures were used to present the cheeses in subsequent SP tasks. Established food labels were used to indicate

[1] During the period of data collection, the USD to NOK exchange rate was approximately USD 1.00 = NOK 6.80.

whether different cheeses were organic, low-fat or low-saturated-fat. The Debio symbol is the official food label for organic products in Norway. The Keyhole symbol is a common Nordic food label that identifies healthier options within specific food groups, such as for example low-fat cheese. The symbol was initiated at the government level and is administered by public health and food agencies. The LHL symbol is owned and administered by the Norwegian Heart and Lung Patient Organization. Against a yearly fee, this organization allows food producers to use the LHL symbol on products that, relative to other products in the same category, may help lower the risk of developing heart diseases. The picture of one of the sixteen cheeses that were used in the experiment is presented in Fig. 2.

Fig. 2. Example of a cheese image that was used in the SP experiment.

The last paragraph in Fig. 1 includes health information and official dietary recommendations concerning the two fat-related cheese attributes. This part of the initial information screen, which serves as a proxy for public provision of diet-related health information, was presented to about half of the participants (n = 208) by random selection.

Following the initial information screen, the participants were asked to state their preferences for different variants of the generic cheese. By random selection and independently of whether they were exposed to the health information on the initial information screen, half of the participants (n = 204) were presented with a series of eight choice sets, one per screen. In each choice set, the participants were asked to choose among two cheeses according to highest probability of buying.[2] The remaining 204 participants were presented with eight cheeses and asked to rank these from the highest to the lowest probability of buying. To ease the cognitive burden of the participants, the ranking was conducted as a series of choices over seven screens. On the first screen, all eight cheeses were shown, and the participants were asked to mark their four most preferred cheeses. The six next screens proceeded as follows. On screen (2), the four selected cheeses from screen (1) were shown, and the participants were asked to click on the most preferred cheese among these (i.e., their top-ranked cheese). On screen (3), the three remaining cheeses from screen (2) were shown, and the participants were asked to click on the most preferred cheese among these. On screen (4), the two remaining cheeses from screen (3) were shown, and the participants were asked to click on the most preferred cheese among these. Screens (5)–(7) proceeded in the same way as screens (2)–(4), but now for the four least preferred cheeses.

The choice experiment was generated using the SAS macro%ChoicEff, which uses a D-optimality algorithm to search for efficient choice designs (Kuhfeld, 2009:764).[3] The attribute combinations of the eight cheeses in the ranking experiment were based on a balanced orthogonal fractional factorial design with four two-level factors. The visual placing of the different cheeses on computer screens and the ordering of the different choice sets was randomized.

[2] The choice sets did not include a 'none-of-these' option, which is sometimes used in choice experiments (Hensher *et al.*, 2005:176).

[3] In our application of %ChoicEff, the full factorial design of $2^4 = 16$ cheeses was used as candidate set. The initial parameters for the four cheese attributes in the binary logit model were set to zero, and no attribute restrictions were applied.

4. Estimation methods

The SP data from the ranking and choice experiments are pooled and analyzed using rank-ordered mixed logit models for panel data (Train, 2003:149, 160). In this specification, for each participant in the ranking experiment, the data is converted into a series of seven choice sets. The first choice set includes all eight cheeses, and the participant 'chooses' the cheese that he or she ranked highest in the experiment. The second choice set includes all eight cheeses minus his or her highest ranked cheese. The third choice set includes all eight cheeses minus his or her two highest ranked cheeses, and so on. The observations from the choice experiment, with two cheeses in each of eight choice sets per participant, are treated as in a standard mixed logit model.

We are mainly interested in analyzing the distributional effects of diet-related health information. Thus, in our main specification of the model, we interact dummy variables for low-saturated-fat (*Lowsat*) and low-fat (*Lowfat*) cheese with education, income, age and gender as defined in Table 1, and with a dummy variable for whether the participant received health information at the beginning of the experiment (*Hinfo*). Pooling the choice and converted ranking data, the utility of cheese j for individual i in choice situation t is:

$$
\begin{aligned}
U_{jit} = \; & \beta_1 Price_{jit} + \beta_{2i} Lowsat_{jit} + \beta_{3i} Lowfat_{jit} + \beta_{4i} Organic_{jit} + \\
& (\beta_5 Lowsat_{jit} + \beta_6 Lowfat_{jit} + \beta_7 Organic_{jit}) \cdot Hinfo_i + \\
& (\beta_8 College_i + \beta_9 Med\text{-}high\ income_i + \beta_{10} Age50\text{–}70_i + \beta_{11} Female_i) \cdot Lowsat_{jit} + \\
& (\beta_{12} College_i + \beta_{13} Med\text{-}high\ income_i + \beta_{14} Age50\text{–}70_i + \beta_{15} Female_i) \cdot Lowsat_{jit} \cdot Hinfo_i + \\
& (\beta_{16} College_i + \beta_{17} Med\text{-}high\ income_i + \beta_{18} Age50\text{–}70_i + \beta_{19} Female_i) \cdot Lowfat_{jit} + \\
& (\beta_{20} College_i + \beta_{21} Med\text{-}high\ income_i + \beta_{22} Age50\text{–}70_i + \beta_{23} Female_i) \cdot Lowfat_{jit} \cdot Hinfo_i + \varepsilon_{jjt}, \quad (1)
\end{aligned}
$$

where *Price* and *Organic* indicate the price of cheese j and whether it was organically produced, respectively, and ε_{jit} is a random error term assumed to be distributed i.i.d. extreme value. The parameters β_2, β_3 and β_4 are assumed to be random parameters that are allowed to vary over individuals, hence the subscript i on these three parameters. The remaining parameters are held fixed. Since it is reasonable to expect that some individuals prefer low-saturated-fat, low-fat and organic cheeses, while others avoid these cheeses, we will assume that β_2, β_3 and β_4 follow an untruncated normal distribution. For these three attributes, the mixed logit model provides estimates of both the mean and the standard deviation of the parameters. The mixed logit models were estimated using the Stata module *mixlogit* (Hole, 2007a), and two thousand Halton draws were used in the simulations. Further details on estimation of mixed logit models are provided in Train (2003) and Hole (2007a).

The parameters of the mixed logit model may be combined to calculate the participants' marginal willingness to pay (MWTP) for low-saturated-fat, low-fat and organic cheese relative to otherwise identical cheeses without these attributes. In the case of an attributes-only model, the mean MWTP for attribute k is calculated as the ratio of β_k to the negative of the price parameter. However, we need to account for the various interaction terms in Eq. (1), and we also multiply the MWTPs by two to obtain MWTPs per kg instead of per 500 g cheese. Thus, as an example, for a participant without college education who received health information at the beginning of the experiment, the MWTP per kg for low-saturated-fat cheese relative to regular-saturated-fat cheese, evaluated at the mean of the other socio-demographic characteristics, is calculated as;

$$\text{MWTP}_{\text{kg }(Lowsat = 1,\, Hinfo = 1,\, College = 0)} = -2 \cdot \left(\frac{\hat{\beta}_2 + \hat{\beta}_5 + \sum_{k=9}^{11} \hat{\beta}_k \bar{\mathrm{x}}_k + \sum_{k=13}^{15} \hat{\beta}_k \bar{\mathrm{x}}_k}{\hat{\beta}_1} \right), \tag{2}$$

where $\bar{\mathrm{x}}_k$ is the mean of the socio-demographic variable corresponding to parameter k in Eq. (1). Confidence intervals for this and similar MWTPs are obtained by combining the relevant parameters and their robust standard errors from the mixed logit model using the delta method (Hole, 2007b).

5. Results

The estimation results from two rank-ordered mixed logit models are reported in Table 2. In Model 1, the overall effects of health information on preferences for low-saturated-fat, low-fat and organic cheese are estimated. This model is given by Eq. (1) where β_8 to β_{23} are restricted to be zero. Model 2 is the unrestricted model given by Eq. (1). In Model 2, we investigate to

Table 2
Rank-ordered mixed logit models for choices of semi-hard cheese.

		Model 1		Model 2	
		Parameter	*t-stat*	Parameter	*t-stat*
Price		-0.082	*-19.12*	-0.082	*-19.12*
Lowsat		0.586	*6.07*	0.316	*1.14*
Lowfat		0.187	*2.56*	0.427	*2.46*
Organic		0.361	*4.39*	0.361	*4.39*
Lowsat×Hinfo		0.426	*2.95*	0.235	*0.61*
Lowfat×Hinfo		0.348	*3.32*	-0.073	*-0.28*
Organic×Hinfo		-0.072	*-0.61*	-0.071	*-0.61*
Lowsat	×College			0.177	*0.78*
	×Med-high income			-0.148	*-0.69*
	×Age 50–70			0.130	*0.69*
	×Female			0.313	*1.63*
Lowsat×Hinfo	×College			-0.372	*-1.20*
	×Med-high income			0.356	*1.16*
	×Age 50–70			0.206	*0.74*
	×Female			0.170	*0.59*
Lowfat	×College			-0.019	*-0.11*
	×Med-high income			-0.045	*-0.29*
	×Age 50–70			-0.293	*-2.10*
	×Female			-0.052	*-0.37*
Lowfat×Hinfo	×College			-0.214	*-0.92*
	×Med-high income			0.152	*0.72*
	×Age 50–70			0.461	*2.29*
	×Female			0.364	*1.81*
Standard deviation parameters					
Lowsat		1.053	*10.28*	1.028	*10.13*
Lowfat		0.484	*3.85*	0.446	*3.41*
Organic		0.724	*6.63*	0.722	*6.58*
Number of choice observations		3,060		3,060	
Number of participants		408		408	
Log likelihood		-2756.97		-2744.30	

Notes: Rank-ordered mixed logit models for choices of semi-hard cheese in the SP experiment, estimated by simulated maximum likelihood. Two thousand Halton draws were used in the simulations. Normal distributions with free variance were assumed for *Lowsat*, *Lowfat* and *Organic*. The models pool choice and converted ranking data.
The *t*-statistics are based on robust standard errors.

what extent the effects of health information on preferences for low-saturated-fat and low-fat cheese vary by education, income, age and gender. Based on a likelihood ratio (LR) test, the restriction in Model 1 of β_8 to β_{23} being zero is rejected at the 90% level, but not at the 95% level (LR test p-value is 0.064).[4]

5.1. Health information – overall effects

The results of Model 1 suggest that on average, participants in both health information groups prefer low-saturated-fat, low-fat and organic cheese. Combining the price parameter and the other parameters as described in Section 4, the mean MWTP for these three cheese attributes in the non-information group are NOK 14.2, NOK 4.5 and NOK 8.8 per kg, respectively (all p-values < 0.01). At the time of the experiment, a 'standard' semi-hard cheese (regular-saturated-fat, regular-fat and conventionally produced) cost between NOK 80–100 per kg. Assuming here a price of NOK 90 per kg, this suggests that on average, participants in the non-information group are willing to pay a price premium of 15.8% for low-saturated-fat cheese, 5.0% for low-fat cheese and 9.8% for organic cheese.[5]

Exposure to health information had a significant effect on preferences for low-saturated-fat and low-fat cheese. This is illustrated in Fig. 3, which is based on the results of Model 1. The light bars show the mean MWTPs for the three cheese attributes in the non-information group, while the dark bars show the corresponding MWTPs in the information group. The mean MWTP for low-saturated-fat and low-fat cheese in the information group are NOK 24.5 and NOK 13.0 per kg, respectively. This is 1.73 and 2.89 times higher than corresponding MWTPs in the non-information group. Preferences for organic cheese were not

[4] The LR test statistic for comparing Model 1 and Model 2 in Table 2 is 25.34. This test statistic is distributed chi-squared with 16 degrees of freedom. The two mixed logit models in Table 2 perform better than corresponding conditional logit models according to similar LR tests (p-values<0.001).

[5] Partly because there is only one provider of low-saturated-fat and organic cheese in the Norwegian cheese market, the real price premiums for these attributes are high, both at around 50%. The price premium for low-fat cheese varies between 0–15%, depending on brand.

significantly affected by exposure to health information. This is not surprising, as the information script did not focus on the properties of organic cheese.

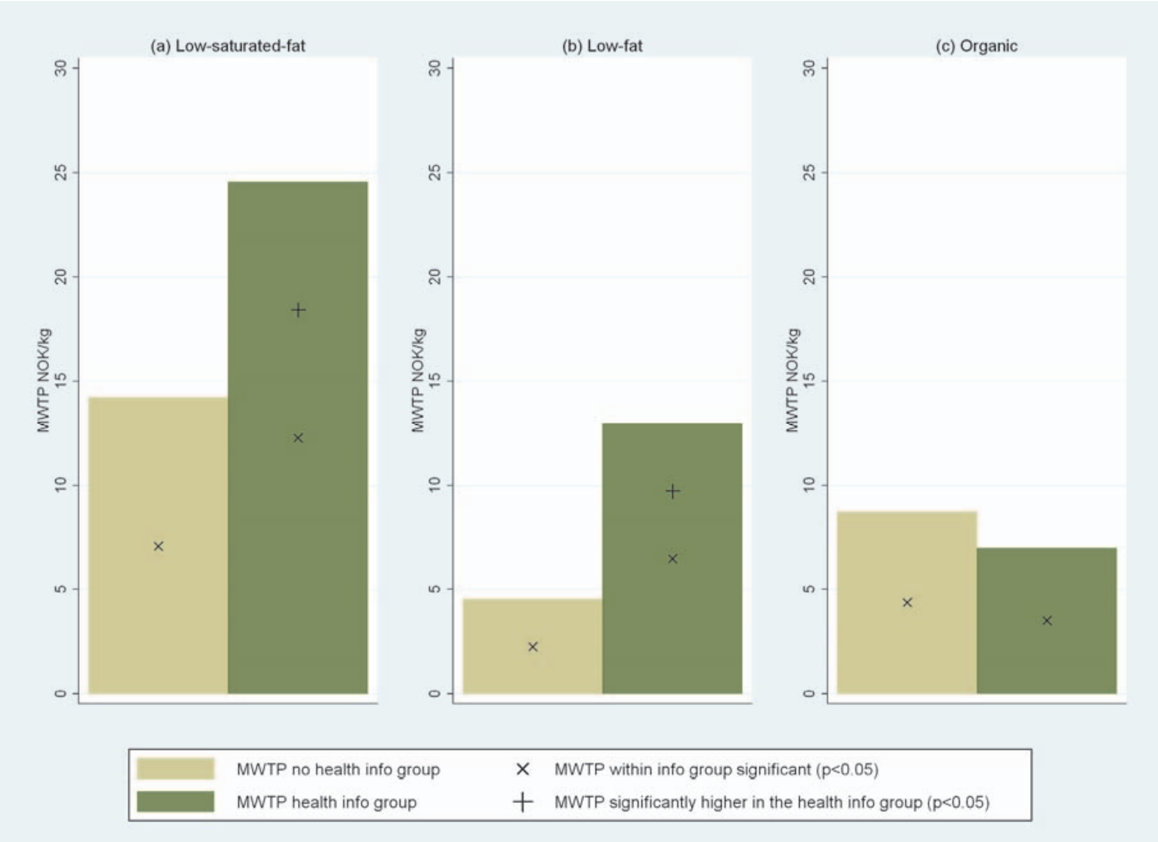


Fig. 3. Mean MWTPs for low-saturated-fat, low-fat and organic cheese, split by the two health information groups. MWTP calculations are based on the results of Model 1 in Table 2. The cost of a 'standard' cheese (regular-saturated-fat, regular-fat and conventionally produced) is about NOK 90 per kg.

The large and statistically significant standard deviation parameters in Model 1 suggest that cheese preferences vary considerably over the participants. While the mean MWTPs for low-saturated-fat, low-fat and organic cheese are positive in both information groups, the standard deviation estimates suggest that these attributes are negatively perceived by respectively 28.9%, 35.0% and 30.9% of the participants in the non-information group, and by 16.8%, 13.4% and 34.5% of the participants in the information group. Thus, these participants are willing to pay a price premium to instead have regular-saturated-fat, regular-fat and conventionally produced cheese.

In both information groups, the mean MWTP for low-saturated fat cheese is considerably higher than the mean MWTP for low-fat and organic cheese. Low-saturated-fat cheese did not exist in the Norwegian cheese market at the time of the experiment but has later been introduced, and so the novelty of this attribute may have attracted many participants. Previous studies have shown that consumers are generally positive to innovations in traditional food products on the condition that sensory quality is maintained (Almli *et al.*, 2011; Guerrero *et al.*, 2009). While low-fat cheese has been on the market for years and is often experienced as less tasty than regular-fat cheese, the low-saturated-fat alternative may have created high sensory expectations in the consumer's mind, as it is not the amount of fat, but only the fat type composition which is modified in this cheese. At the same time, taste considerations may explain why some participants are willing to pay a price premium to avoid having low-saturated fat and low-fat cheese.

5.2. Health information – distributional effects

Based on the results of Model 2 in Table 2, we have estimated mean MWTPs for low-saturated-fat and low-fat cheese split by (i) the two health information groups and (ii) the two education, income, age and gender groups. These results are presented in Fig. 4. Point estimates and *t*-statistics for the different MWTPs in Fig. 4, as well as for MWTP differences between different groups of participants, are reported in Table 3.

Thirty-two group-attribute specific MWTPs are shown in Fig. 4. Out of these, only five are not statistically significant at the 95% level, as indicated by absent × symbols inside the vertical MWTP bars. These five insignificant MWTPs relate to low-fat cheese and groups of participants who were not exposed to health information in the experiment.

The effects of health information on additional MWTP for low-saturated-fat and low-fat cheese are larger among non-college, medium-high income, age 50–70 and female participants than among their respective counterparts, i.e., college-educated, low income, age 30–49 and male participants. In the former groups, the information effects are also always

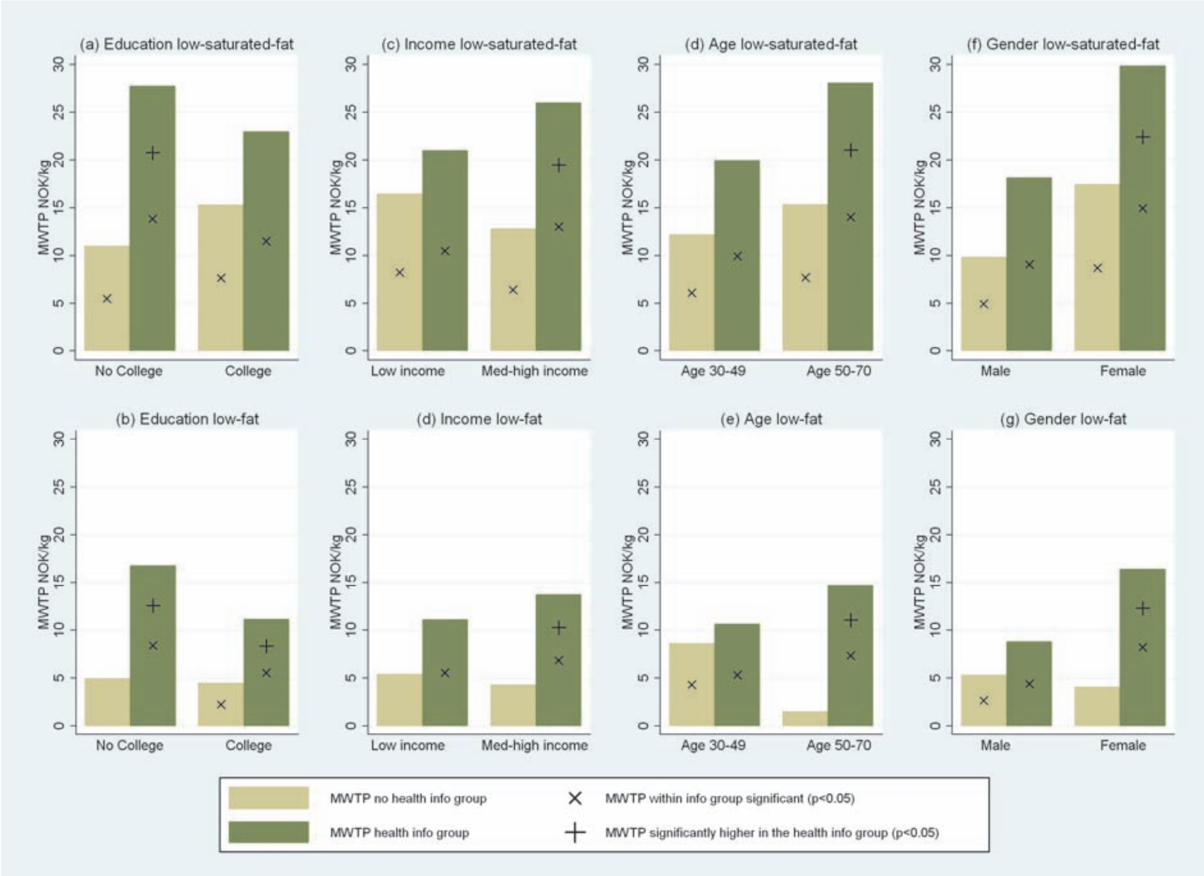


Fig. 4. Mean MWTPs for low-saturated-fat and low-fat cheese, split by the two health information groups and socio-demographic characteristics. MWTP calculations are based on the results of Model 2 in Table 2. The cost of a 'standard' cheese (regular-saturated-fat, regular-fat and conventionally produced) is about NOK 90 per kg.

Table 3
Group-specific MWTPs for low-saturated-fat and low-fat cheese.

	MWTP low-saturated-fat cheese (NOK/kg)						MWTP low-fat cheese (NOK/kg)					
	(1)		(2)		(3)=(2)-(1)		(4)		(5)		(6)=(5)-(4)	
	HI=0	*t-stat*	HI=1	*t-stat*	HI eff.	*t-stat*	HI=0	*t-stat*	HI=1	*t-stat*	HI eff.	*t-stat*
Non-college	11.01	*2.30*	27.73	*6.72*	16.72	*2.66*	4.96	*1.43*	16.78	*4.99*	11.82	*2.45*
College	15.30	*5.80*	22.98	*7.01*	7.69	*1.83*	4.50	*2.19*	11.13	*5.35*	6.63	*2.26*
Difference in HI effect					**-9.03**	***-1.20***					**-5.19**	***-0.92***
Low income	16.45	*3.80*	20.98	*4.96*	4.53	*0.76*	5.40	*1.85*	11.10	*4.16*	5.70	*1.45*
Med-high income	12.85	*4.56*	26.01	*7.94*	13.16	*3.03*	4.32	*1.95*	13.70	*5.96*	9.39	*2.92*
Difference in HI effect					**8.63**	***1.16***					**3.69**	***0.72***
Age 30–49	12.18	*3.56*	19.92	*6.53*	7.74	*1.68*	8.62	*3.56*	10.62	*4.36*	2.00	*0.59*
Age 50–70	15.33	*4.89*	28.06	*6.98*	12.73	*2.53*	1.52	*0.63*	14.71	*5.74*	13.19	*3.72*
Difference in HI effect					**4.99**	***0.74***					**11.19**	***2.29***
Male	9.86	*3.14*	18.18	*5.47*	8.32	*1.83*	5.32	*2.56*	8.83	*3.82*	3.50	*1.13*
Female	17.45	*5.12*	29.88	*7.53*	12.43	*2.38*	4.06	*1.52*	16.41	*6.17*	12.34	*3.25*
Difference in HI effect					**4.11**	***0.59***					**8.84**	***1.81***

Notes: MWTPs for low-saturated-fat and low-fat cheese, split by the two health information groups and socio-demographic characteristics. Calculations based on the results of Model 2 in Table 2 using the delta method. HI = 1 if participants received additional diet-related information and HI = 0 otherwise.

statistically significant at the 95% level, as indicated by the + symbols inside the dark MWTP bars in Fig. 4. They are also generally quite substantial. For example, among non-college participants who received health information, the mean MWTP for low-saturated-fat and low-fat cheese are NOK 27.7 and NOK 16.8 per kg, respectively. These values are 2.52 and 3.36 times higher than corresponding MWTPs for non-college participants who did not receive health information (NOK 11.0 and NOK 5.0 per kg, respectively).

Although smaller, the information effects are always positive also among college-educated, low income, age 30–49 and male participants. However, except for the effect of information on additional MWTP for low-fat cheese among college-educated participants (Fig. 4b), these information effects are not statistically significant at the 95% level. Statistical tests for differences in information effects within each socio-demographic characteristic are reported in Table 3 (in bold). Only the effect of information on additional MWTP for low-fat cheese among age 50–70 participants *minus* the corresponding information effect among age 30–49 participants is statistically significant at the 95% level. Differences between non-college and college, low and medium-high income, and male and female participants are insignificant.[6]

5.3. Possible mechanisms – the role health knowledge, awareness and prices

Following the SP experiment, the participants were presented with a number of Likert-type statements on various diet and health issues. Some possible explanations for the above results may be found by exploring the responses to these statements. We have estimated ordered logit models (Greene, 2003) for six of the statements, where each statement is regressed on

[6] To check for robustness of the results in Model 2, we have estimated several alternative versions of this model, including the standard conditional logit model and a mixed logit model where we use continuous age and income variables and then evaluate MWTPs in the two health information groups at age 35 and age 65 and at the 25th and the 75th percentile of the income distribution. The main results are generally very similar across Model 2 and these and other alternative model specifications. One notable exception is that when we control for continuous age and income variables, the larger effect of information on preferences for low-fat cheese among female relative to male participants becomes statistically significant at the 95% level.

education, income, age and gender as defined in Table 1. The results of these models as well as the statements themselves are reported in Table 4. For purposes of interpretation, we also report the marginal effects (ME), within each statement, on the probability of choosing the highest response alternative (e.g., 'totally agree') relative to choosing one of the other six or four response alternatives.

Table 4
Subjective statements on food preferences – results from odered logit models.

Statements
S1: Price is important to me when I decide what to eat on regular weekdays
S2: Saturated fat content is important to me when I decide what to eat on regular weekdays
S3: Fat content is important to me when I decide what to eat on regular weekdays
S4: Staying healthy is important to me when I decide what to eat on regular weekdays
S5: Total calories is important to me when I decide what to eat on regular weekdays
S6: I am well informed about the long-term associations between diet and health

	S1		S2		S3		S4		S5		S6	
Variable	β	*ME*	β	*ME*	β	*ME*	β	*ME*	β	*ME*	β	*ME*
College	-0.41*	*-0.03*	-0.04	*-0.01*	-0.09	*-0.01*	0.16	*0.03*	-0.27	*-0.03*	0.79*	*0.16*
Med-high income	-0.94*	*-0.09*	-0.12	*-0.01*	0.21	*0.02*	0.00	*0.00*	-0.05	*0.00*	0.02	*0.00*
Age 50–70	-0.28	*-0.02*	0.69*	*0.08*	0.25	*0.02*	1.06*	*0.21*	0.36*	*0.03*	-0.07	*-0.01*
Female	-0.34	*-0.03*	0.53*	*0.06*	0.38*	*0.03*	0.68*	*0.14*	0.61*	*0.06*	0.48*	*0.10*
μ_1	-4.65*		-3.43*		-2.93*		-3.72*		-2.65*		-3.52*	
μ_2	-3.41*		-2.48*		-1.95*		-3.31*		-2.01*		-3.11*	
μ_3	-2.54*		-1.49*		-1.03*		-2.89*		-1.20*		-1.59*	
μ_4	-1.40*		-0.16		-0.17		-1.03*		0.18		1.48*	
μ_5	-0.21		1.10*		1.24*		0.63*		1.57*			
μ_6	1.07*		2.49*		2.68*		1.95*		2.49*			
Log-likelihood	-713.75		-670.49		-700.21		-555.34		-693.46		-386.69	
No. of obs	408		408		408		408		408		407	

Notes: Statements S1–S5 include seven response alternatives ranging from 'totally disagree' to 'totally agree'. Statement S6 includes five response alternatives ranging from 'very uninformed' to 'very well informed'. Statistically significant parameters at the 95% level are marked * (based on robust standard errors). μ_1–μ_6 are the estimated cut-off points in the ordered logit models. The numbers in *italics* are the marginal effects, within each statement, on the probability of choosing the highest response alternative (e.g., 'totally agree') relative to choosing one of the other six (S1–S5) or four (S6) response alternatives.

Four of the six statements in Table 4 relate to aspects of diet-health awareness (S2–S5). Results of the ordered logit models show that these statements are closely associated with age and gender, but not with education and income. When deciding what to eat, age 50–70 and female respondents are relatively more concerned than age 30–49 and male respondents about saturated fat (S2) and total fat contents (S3), amounts of calories consumed (S5) and in particular the role of diets in staying healthy (S4). Such awareness patterns for the two fat-related statements (S2 and S3) have also been found in the US by Variyam (1999). These results correspond logically to the results in the cheese experiment, where age 50–70 and female participants were more clearly affected by health information than age 30–49 and male participants. Several factors may explain these differences. For example, the risk of developing heart diseases and fat-related cancers are higher at later stages of the life course, and thus older people are relatively more likely than younger people to care about health issues and show interest for diet-related health information. Furthermore, results from experimental studies suggest that women are generally more risk averse than men (Croson and Gneezy, 2009). Our results may reflect this phenomenon in the case of health risks.

The statement on the participants' general diet-health knowledge (S6) is strongly correlated with gender and education, but not with age and income. Respondents who have attended college or university are sixteen percentage points more likely than respondents with less education to state that they are 'very knowledgeable' about the long-term associations between diet and health. This finding corroborates well with the results in the cheese experiment; participants without college education were more clearly affected by health information than college-educated participants, and this seems to reflect different *a priori* levels of diet-health knowledge, as indicated by the results in Table 4. The results for gender show that men are both less knowledgeable about diet-health associations than women (Table 4) and less responsive to diet-related health information in the SP experiment (Fig. 4f–g). As noted, the statements on diet-health awareness (S2–S5) are strongly correlated with gender, but somewhat surprisingly, not with education. Thus, while insufficient diet-health knowledge (S6) may represent a constraint for making healthful diet choices among both non-college and male participants, only male participants are constrained by a general lack of diet-health awareness. According to our results, in males, this last constraint seems to dominate the former.

The first statement in Table 4, the importance of food prices (S1), is significantly associated with education and in particular income, as participants in the medium-high income group are nine percentage points less likely than participants in the low income group to ‘totally agree’ in the that prices are important when buying foods. At the same time, income is not significantly associated with diet-health knowledge and awareness (S2–S6). Thus, our finding that medium-high income participants are more clearly affected by health information than low income participants seems to mainly reflect the fact that the information effects are measured in terms of marginal willingness to pay, which is likely to depend in part on income and associated budget constraints for food.

6. Discussion and conclusion

Our daily decisions about eating healthy or unhealthy foods are influenced by a highly complex mix of factors. Nutrition policies may target at least two of these factors, health knowledge and awareness, through dissemination of diet-related health information, using for example media campaigns. It is difficult to isolate the effects of such health information policies using non-experimental data, including their distribution across different socio-demographic groups. To investigate the effects of health information, we have examined these issues using experimental data. Our stated preference experiment focused on healthy attributes in a generic everyday use semi-hard cheese. Half of the participants were exposed to objective health information related to cheese consumption prior to performing either a choice or a ranking task.

Our results show that preferences for low-saturated-fat and low-fat cheese are strongly affected by exposure to health information. Education is found to be a strong indicator of prior diet-health knowledge but is simultaneously unrelated to subjective indicators of diet-health awareness. In our cheese experiment, non-college participants are therefore more clearly affected by health information than college-educated participants. Although our results suggest a promising role for health information policies in reducing educational differences in diet-health knowledge and thus dietary behavior, a challenge remains in how to effectively target low education groups in non-experimental settings. Also, according to our results, targeting low income groups, young people and in particular males through health information policies seems difficult. Rather than providing generic health information, experiences from smoking suggest that using ads or campaigns that contain personal stories or highly emotional elements such as films and images showing blocked blood vessels, tumors, heart attacks, and so on that could result from years of cigarette smoking may be efficient in reaching young people and low socioeconomic status groups (Durkin *et al.*, 2009).

Current leading nutrition information initiatives such as 'MyPlate' in the US, the similar 'Eatwell Plate' in the UK, the 'Keyhole' symbol in the Nordic countries, the '5 A Day' campaign in various European countries and official dietary guidelines tend to focus on well-balanced diets and the identification of healthy dietary choices. While this is expected to educate people and thus have positive overall effects on dietary behavior, it would be interesting, at least as a research exercise, to compare the effects of such positively loaded nutrition messages with the effects of more emotional and negatively loaded messages such as those described above for smoking, including their distribution across different socio-demographic groups. While it may be more attractive to encourage people to eat healthy than to scare them from eating unhealthy, these two approaches – positively versus negatively loaded health information messages – should be viewed in light of their likely effectiveness, including their ability to reduce socioeconomic inequalities in dietary behavior and improve overall population health, and thereby reduce the direct and indirect societal costs that are associated with obesity and diet-related chronic diseases.

This study uses a non-representative sample and possibly suffers from various forms of hypothetical bias (Hensher, 2010). Moreover, it focuses only on the consumption of everyday use semi-hard cheese. More studies that examine how diet choices are affected by exposure to health information, including to what extent such effects vary by socio-demographic characteristics, are therefore needed. Relevant extensions include studies on other food items, studies from other countries, the use of non-hypothetical settings such as field experiments with binding choices, assessing the duration of information effects via follow-ups, and testing different types of health message formats, including positive and negative information.

References

Alfnes F, Rickertsen K. 2003. European consumers' willingness to pay for US beef in experimental auction markets. *American Journal of Agricultural Economics* **85:** 396–405.

Almli VL, Næs T, Sulmont-Rossé C, Enderli G, Issanchou S, Hersleth M. 2011. Consumers' acceptance of innovations in traditional cheese. A comparative study in France and Norway. *Appetite* **57**: 110–120.

Balcombe K, Fraser I, Di Falco S. 2010. Traffic lights and food choice: a choice experiment examining the relationship between nutritional food labels and price. *Food Policy* **35**: 211–220.

Blaylock J, Smallwood D, Kassel K, Variyam J, Aldrich L. 1999. Economics, food choices and nutrition. *Food Policy* **24**: 269–286.

Camerer C, Loewenstein G, Prelec D. 2005. Neuroeconomics: how neuroscience can inform economics. *Journal of Economic Literature* **43**: 9–64.

Croson R, Gneezy U. 2009. Gender differences in preferences. *Journal of Economic Literature* **47**: 448–474.

CSDH. 2008. Closing the gap in one generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health. World Health Organization: Geneva.

Durkin S, Biener L, Wakefield M. 2009. Effects of different types of antismoking ads on reducing disparities in smoking cessation among socioeconomic subgroups. *American Journal of Public Health* **99**: 2217–2223.

Etilé F. 2007. Social norms, ideal body weight and food attitudes. *Health Economics* **16**: 945–966.

Gracia A, Loureiro M, Nayga Jr. RM. 2009. Consumers' valuation of nutritional information: a choice experiment study. *Food Quality and Preference* **20**: 463–471.

Greene WH. 2003. *Econometric Analysis* (5th ed.). Prentice Hall: Upper Saddle River, NJ.

Grossman M. 2000. The human capital model of the demand for health. In *Handbook of health economics*. Culyer AJ, Newhouse JP (Eds.). Elsevier: Amsterdam.

Guerrero L, Guardia MD, Xicola J, Verbeke W, Vanhonacker F, Zakowska-Biemans S, *et al.* 2009. Consumer-driven definition of traditional food products and innovation in traditional foods. A qualitative cross-cultural study. *Appetite* **52**: 345–354.

Hensher D, Rose JM, Greene WH. 2005. *Applied choice analysis: a primer*. Cambridge University Press: Cambridge.

Hensher DA. 2010. Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B* **44**: 735–752.

Hole AR. 2007a. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* **7**: 388–401.

Hole AR. 2007b. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics* **16**: 828–840.

Hossain P, Kawar B, El Nahas M. 2007. Obesity and diabetes in the developing world: a growing challenge. *New England Journal of Medicine* **356**: 213–215.

Huffman W, Rousu M, Shogren JF, Tegene A. 2007. The effects of prior beliefs and learning on consumers' acceptance of genetically modified foods. *Journal of Economic Behavior and Organization* **63**: 193–206.

Ippolito PM, Mathios AD. 1994. Information, policy, and the sources of fat and cholesterol in the U.S. diet. *Journal of Public Policy & Marketing* **13**: 200–217.

Kuhfeld WF. 2009. *Marketing research methods in SAS*. SAS Institute Inc.: Cary.

Loewenstein G, Prelec D. 1992. Anomalies in intertemporal choice: evidence and an interpretation. *Quarterly Journal of Economics* **107**: 573–597.

Lusk JL, Fields D, Prevatt W. 2008. An incentive compatible conjoint ranking mechanism. *American Journal of Agricultural Economics* **90**: 487–498.

Nayga Jr. RM. 2008. Nutrition, obesity, and health: policies and economic research challenges. *European Review of Agricultural Economics* **35**: 281–302.

Niederdeppe J, Kuang X, Crock B, Skelton A. 2008. Media campaigns to promote smoking cessation among socioeconomically disadvantaged populations: what do we know, what do we need to learn, and what should we do now? *Social Science & Medicine* **67**: 1343–1355.

Norwegian Directorate of Health. 2010. Utviklingen i norsk kosthold. Matforsyningsstatistikk og forbruksundersøkelser. Helsedirektoratet IS–1873: Oslo.

Norwegian Directorate of Health. 2011. Kostråd for å fremme folkehelsen og forebygge kroniske sykdommer. Metodologi og vitenskaplig kunnskapsgrunnlag. Nasjonalt råd for ernæring. Helsedirektoratet IS–1881: Oslo.

Norwegian Ministry of Health and Care Services. 2006. National strategy to reduce social inequalities in health*.* *Report No. 20 (2006-2007) to the Storting*: Oslo.

OECD. 2008. *Growing unequal? Income distribution and poverty in OECD countries*. OECD Publishing: Paris.

Park J, Davis GC. 2001. The theory and econometrics of health information in cross-sectional nutrient demand analysis. *American Journal of Agricultural Economics* **83**: 840–851.

Roosen J, Marette S. 2011. Making the 'right' choice based on experiments. *European Review of Agricultural Economics* **38**: 361–381.

Roosen J, Marette S, Blanchemanche S, Verger P. 2009. Does health information matter for modifying consumption? A field experiment measuring the impact of risk information on fish consumption. *Review of Agricultural Economics* **31**: 2–20.

Train K. 2003. *Discrete choice methods with simulation*. Cambridge University Press: Cambridge.

Variyam JN. 1999. Role of demographics, knowledge and attitudes. In *America's eating habits: changes & consequences*. Frazao E (Ed.). United States Department of Agriculture, Economic Research Service: Washington, DC.

Variyam JN. 2008. Do nutrition labels improve dietary outcomes? *Health Economics* **17**: 695–708.

Wakefield MA, Loken B, Hornik RC. 2010. Use of mass media campaigns to change health behaviour. *Lancet* **376**: 1261–1271.

Wardle J, Parmenter K, Waller J. 2000. Nutrition knowledge and food intake. *Appetite* **34**: 269–275.

Wisdom J, Downs JS, Loewenstein G. 2010. Promoting healthy choices: information versus convenience. *American Economic Journal: Applied Economics* **2**: 164–178.

World Health Organization. 2003. Diet, nutrition and the prevention of chronic diseases. *WHO Technical Report Series No. 916*. WHO: Geneva.

Arnstein Øvrum

Arnstein Øvrum was born in Skien, Norway, in 1977. He holds a MSc. Degree in Development and Resource Economics from the Norwegian University of Life Sciences (2006). Arnstein is currently employed at the Norwegian Agricultural Economics Research Institute. The thesis focuses on the relationship between socioeconomic status, health and related lifestyle choices. It uses repeated cross-section and stated preference data from Norway and consists of an introduction and four independent papers. Paper 1 compares sources of inequality in health, represented by self-assessed health and obesity, with sources of inequality in lifestyle choices central to the production of health, represented by physical activity, cigarette smoking and two indicators of healthy dietary behavior; the consumption of fish and the consumption of fruits and vegetables. The results demonstrate that patterns of inequality in health are not necessarily representative of patterns of inequality in important, underlying production factors of health. Paper 2 examines how education and income differences in physical activity, the consumption of fruits and vegetables, cigarette smoking and self-assessed health evolve over the adult life course. Although mixed, the results provide some evidence of increased health consciousness and associated lifestyle improvements in age among lower socioeconomic status groups. Such improvements may potentially contribute to reducing cumulative advantage effects in health by socioeconomic status at older ages. Paper 3 estimates the demand for physical activity and fruits and vegetables using latent class models, focusing on subpopulation heterogeneity in the effects of education and income. The results suggest that among the majority of the population that should be more physically active and eat more fruits and vegetables, the role of education and income may be even more important than previously assumed. Paper 4 uses stated preference data on semi-hard cheese to examine how diet choices are affected by exposure to health information. The results suggest a promising role for health information policies in reducing educational differences in diet-health knowledge and thus dietary behavior. Targeting low income groups, young people and particularly males through health information policies seems more difficult.

Professor Kyrre Rickertsen was Arnstein's main supervisor. Associate Professor Frode Alfnes was co-supervisor.

E-mail:arnstein.ovrum@nilf.no

Department of Economics and Resource Management
Norwegian University of Life Sciences
PO Box 5003
N-1432 Ås,
Norway

Telephone: +47 6496 5700
Telefax: +47 6496 5701
e-mail: *ior@umb.no*
http:/www.umb.no/ior

ISSN: 1503-1667
ISBN: 978-82-575-1042-8