DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 50, ARTICLE 25, PAGES 667–732
PUBLISHED 5 APRIL 2024
http://www.demographic-research.org/Volumes/Vol50/25/
DOI: 10.4054/DemRes.2024.50.25

Research Article

The importance of correcting for health-related survey non-response when estimating health expectancies: Evidence from the HUNT Study

Fred Schroyen

Contents

Demographic Research: Volume 50, Article 25
Research Article

The importance of correcting for health-related survey non-response when estimating health expectancies: Evidence from the HUNT Study

Fred Schroyen[1]

Abstract

BACKGROUND
Most studies on health expectancies rely on self-reported health from surveys to measure the prevalence of disabilities or ill health in a population. At best, such studies only correct for sample selection based on a limited number of characteristics observed on the invitees.

OBJECTIVE
Using longitudinal data from the Trøndelag Health Study (HUNT), I investigate the extent to which adjustments for a health-related sample selection affect the age profiles for the prevalence of functional impairment (FI) and the associated disability-free life expectancy (DFLE).

METHODS
I estimate a probit model with sample selection under the identifying restriction that the strength of the health-related selection is of similar order to the strength of the selection on observable characteristics. I then compute the selection-adjusted FI prevalence rates and trace out the implications for DFLE using the Sullivan method.

RESULTS
The analysis confirms that poor health measured at younger ages correlates with non-response behaviour in later waves of the survey, and that even for a conservative lower bound for the assumed degree of health-related selection, the estimated age profiles for DFLE lie systematically below the corresponding profiles when controlling only for selection on observable characteristics.

CONCLUSIONS
Health related non-response downwardly biases the raw sample prevalence rates for FI obtained from survey data and contributes to overestimating the expansion in DFLE.

[1] Department of Economics, NHH Norwegian School of Economics, Bergen, Norway.
Email: fred.schroyen@nhh.no

CONTRIBUTION
I present a statistical framework for taking health-related survey non-responses into account when estimating the prevalence rate of FI. The framework can be used to gauge the sensitivity of estimated (changes in) DFLE to health-related sample selection.

1. Introduction

In many developed countries, the responsibility for financing health services and care at older ages, and in some countries also the provision of such services, is assumed by the public sector. Increased life expectancy (LE) may in this respect pose a concern to the extent that the expected period of malfunctioning and need for care grow at the same pace as the expected total lifetime does. This has prompted the need for reliable summary measures of population health and disabilities. These are referred to as health expectancies.

Health expectancy measures have been developed since the early 1970s, and multinational organisations, such as the World Bank, the OECD, and the European Union, as well as many individual countries are regularly reporting them and using them as a tool for monitoring population health and for making informed health policy choices (Saito, Robine, and Crimmins 2014, European Commission 2018). The typology of health expectancies has changed over the years (Mathers 2002: 181), and several authors have proposed different classifications. Robine (2002) proposes to distinguish between health-state expectancy (HSE) indicators, which measure the expected lifetime that a person spends in a particular health state, and health-adjusted life expectancy (HALE) indicators, which provide a summary measure of population health by transforming expected years spent in particular health states into equivalent years of full health. Loosely speaking, with n (≥ 2) mutually exclusive and exhaustive health states, the corresponding HSEs should add up to LE. A HALE measure, however, will first multiply each less-than-full health state expectancy with a weight less than one and then add up these weighted expectancies. A particular measure that does this is the disability-adjusted life expectancy (Mathers 2002), which was used in the WHO project on measuring the global burden of disease.[2]

In some studies, the partitioning of health states is very coarse, (e.g., by distinguishing only between a state with and a state without functional impairment), whereas other

[2] In addition to these descriptive indicators, researchers have also developed health gap indicators. These are normative indicators measuring the distance between a population's current health and a pre-determined normative or ideal level. Disability-adjusted life years (aka DALYs) is the prominent indicator in this class, also developed in connection with the Global Burden of Disease project (Lamb and Siegel 2004: 361; Murray and Lopez 1996; Mathers et al. 2000).

studies go further, (e.g., by also distinguishing between moderate and severe impairment, or by employing a finer partitioning based on the WHO International Classification of Functioning, Disability and Health). Further distinctions can be made according to the type of instrument used to identify disability. The healthy life years indicator of the EU is based on the global activity limitation indicator (GALI) used in the EU-SILC survey.[3] Other indicators are based on the number of activities of daily living or instrumental activities of daily living that a person can perform (De Carvalho Yokota and Van Oyen 2020).

In this article, I use an instrument similar to the GALI, inquiring about the existence of functional impairment (FI)[4] in everyday life due to a long-term illness or injury, to compute functionally impaired life expectancy ($FILE$) and its complement disability-free life expectancy ($DFLE$). Thus $LE = FILE + DFLE$.

Population health indicators have also been prominent in shedding light on an academic debate, starting in the late 1970s, on whether the period of morbidity is becoming more compressed during an average lifetime, or whether it is expanding as expected lifetime increases. The optimistic stance of compression (Fries 1980, 2003) is based on the assumption of a fixed upper limit to the lifespan and contends that medical progress postpones the onset of chronic and irreversible illnesses towards the end of life. The expansion view (Gruenberg 1977; Kramer 1980) argues that the drop in old-age mortality because of life-extending medical advances leads to more years of suffering from chronic disabling diseases without a significant impact on the age of onset of these diseases. In addition, increases in survival rates mean that people are more exposed to the risk of acquiring disabling chronic diseases. These forces imply that the average person spends a larger share of his or her lifetime in disability. An intermediate view, called dynamic equilibrium, which has also received attention and empirical support (Manton 1982), argues that the reduction in mortality risk is correlated with the reduction in the exposure to and severity of chronic disabling diseases, and therefore that increased longevity is accompanied by increased time both in good and poor health.[5]

Irrespective of whether a population health indicator is used for public health planning purposes or for shedding light on the compression–expansion of morbidity debate, it is clear that its statistical properties should be of high quality. The accurate measurement of $DFLE$ and related measures depends on reliable data about two risks occurring to a population: the risk of death and the risk of FI (or disability or ill-health). Several methods exist to measure $DFLE$, the simplest being the observed prevalence life table,

[3] The GALI was proposed in 2001. It is based on the answer to the question "For at least the past six months, to what extent have you been limited because of a health problem in activities people usually do?" with possible answers "Severely limited," "Limited but not severely," and "Not limited at all" (De Carvalho Yokota and Van Oyen, 2020:10).

[4] Given the frequent use of the term "functional impairment" in this article, I abbreviate it as FI (not italicised). Later, in Section 3, I define a specific indicator variable measuring functional impairment. This variable will be denoted as FI (italicised).

[5] See the discussion in Cai and Lubitz (2007) and the recent survey by Robine et al. (2020).

also known as the Sullivan method (Sullivan 1971). This method combines mortality data from civil registries with prevalence data about health status from a cross-section of the population. When the incidence of ill health is stable over time or evolves in a gradual way, the evolution of period prevalence can be tracked by the observed prevalence.[6] In contrast to information about mortality, which is often well registered for entire populations and available over time for countries globally, population measures of health status are rarely available. Instead, researchers rely on survey information about health collected from respondents sampled from the population. Unfortunately, like many other surveys, health surveys also face the problem of non-response. Although survey designs often take non-participation into account by oversampling certain parts of the population, the representativeness of the sample cannot be taken for granted. Response rates in health surveys vary considerably across countries (Jones, Koolman, and Rice 2006; Mindell, Giampaoli, and Goesswald 2015), and rising non-response rates are evident in the most widely used health surveys in the United States (Meyer, Mok, and Sullivan 2015; Czajka and Beyler 2016), in the United Kingdom (Contoyannis, Jones, and Rice 2004; Matthews et al. 2013), and in many other countries (Mindell, Giampaoli, and Goesswald 2015).

The potential bias that survey non-response may cause in the measurement of health status is only to a limited extent addressed in empirical research on changes in $DFLE$. The most common approach is to weight survey responses by sampling characteristics available from the sampling frame, such as gender, age, and region of residence. If surveys are longitudinal, selection into second or later waves can also be analysed and adjusted for by weights obtained from information provided in earlier waves, including health (Jones et al., 2006). These approaches go under the name inverse-probability weighting. If certain observable characteristics of the invited subjects are associated with the probability with which these subjects decline to participate, then one can give extra sample weight to participants with those characteristics to account for the many others who do not accept the invitation. Such inverse-probability weighting corrects for selection into the sample being dependent on observable characteristics – in statistical jargon, when the non-responding individual is 'missing at random.' Recently, some surveys have started to collect para-data – auxiliary variables, such as observations by the interviewer on respondents' willingness to participate and/or proxy measures of key outcome variables – as candidates for non-response adjustments. An assessment of the performance of such variables as a means to adjust for non-response bias concludes that, in general, their correlation with response propensities and relevant outcome variables is weak (Kreuter et al. 2010).

The reason why non-response in health surveys is particularly challenging and why inverse-probability weighting may fail to correct for it is that the object that one tries to

[6] See Cambois, Robine, and Brouard (1999). These authors also describe two other methods – the multiple-decrement life tables and the increment-decrement life tables – which impose stronger data requirements. The issue of sample selection that I discuss in terms of the prevalence life table method applies a fortiori to these other approaches.

measure – health status – may simultaneously influence the ability and/or willingness to participate in the survey. This will be particularly the case for health surveys that involve objective measurement of health conditions because such surveys often require participants to attend a health check facility for anthropometric measuring and collection of blood and urine samples. In the statistical and econometric literature, sample selection is then coined endogenous: Selection into the sample then also depends on unobservable characteristics that are correlated with unobservable characteristics affecting the outcome variable of interest (health). In that case, non-response is no longer 'missing at random.' To my knowledge, endogenous selection, though sometimes acknowledged, has not been formally addressed in the literature on health expectancies.[7]

The standard solution to the endogenous selection problem is to jointly model and estimate the participation decision and the outcome variable of interest in a (semi-) parametric bivariate framework.[8] A common feature of such models is their reliance on a so-called instrument: a variable that shifts the probability of participation but that can be convincingly argued not to affect the outcome variable (health) – namely, to be excluded from the outcome equation. This is necessary to estimate the model reliably. Unfortunately, health surveys rarely include such a variable.

To circumvent such an exclusion restriction, I follow the idea developed by Altonji, Elder, and Taber (2005) that under certain conditions the degree of selection on observables can serve as a guide to the degree of selection on unobservables. This idea is based on the observation that for many large-scale, multi-purpose surveys, background variables that are associated with a large set of outcome variables are more likely to be included in the survey, whereas other variables are omitted because their measurement is too costly, not sufficiently precise, or has too little general relevance. From the perspective of one particular outcome variable of interest, the set of available explanatory variables can then be regarded as a random selection from the total set of variables that explain that outcome variable. As a result, the strength of selection into the survey due to unobservables should be of a similar order as the strength of selection on observables.

In this article, I estimate a bivariate probit model for participation and FI using the relation between the two selection strengths as an identifying restriction. I use data from a large-scale and multi-purpose Norwegian longitudinal health survey: the Trøndelag Health Study (HUNT), collected in three waves from around 1985 to around 2007. In each wave, all residents from the Nord-Trøndelag province ($\sim 90,000$) were invited for a health check. Because of the high response rate in the first wave ($\geq 90\%$), and all participants'

[7] An electronic search through the recently published 300 page volume *International Handbook of Health Expectancies* (Jagger et al. 2020) gives only five hits for 'non-participation' and two hits for 'attrition.'

[8] The original selection model of Heckman (1976, 1979) was for a continuous outcome variable, and the estimation of that model involved a two-step method – at that time the only practical way of estimation. Adaptation of the selection model to a binary outcome variable and maximum likelihood estimation of that model was proposed by Van de Ven and Van Praag (1981). De Luca and Perotti (2011) develop a method to estimate a semi-parametric specification of the latter model. Greene (2018: ch 19) provides an overview.

initial consent to having their data merged with longitudinal population registers, I have exceptionally rich data on background characteristics, labour force participation, and income for prime-aged and elderly responders as well as non-responders. I combine this information with information about response behaviour and initial health status to control for non-response bias in estimates of age- and gender-specific FI prevalence rates in the third wave 22 years later. In a next step, I make use of the Sullivan method to map the age profiles for $FILE$ and its complement $DFLE$.

During the period of observation, life expectancy for Norwegian men (women) aged 50 to 54 years increased by 4.2 (2.5) years. Because certified life tables for Nord-Trøndelag do not exist, I use the ones for the entire country as a substitute.[9] Without correcting for non-response bias, estimates of the prevalence of FI suggest that $DFLE$ increased by 3.3 years for men aged 50 to 54 years and by 1.7 years for women. However, when I correct for selection, I find only a 1.4 year increase in $DFLE$ for men and a 1.6 year reduction for women.

Although I do not claim that the estimates of $DFLE$ are totally unbiased, nor that the assumption about equal selection strength cannot be rejected if the necessary data were available (I will actually conjecture that selection on unobservables is likely to be stronger than selection on observables), the analysis provides strong evidence that health-related non-response seriously biases the estimates for $DFLE$ and related concepts. Invited persons with health limitations are less prone to participate in the HUNT studies than are healthy persons. Ignoring this in the estimation of the population prevalence of FI produces downwardly biased estimates for the $FILE$. Since any life expectancy measure for age a is based on the person-years lived beyond that age, biases in prevalence rates beyond that age accumulate and may therefore produce substantial biases in $FILE$ at age a. Based on imperfect but nevertheless more plausible identifying assumptions about selection bias, my contribution shows that health-related survey non-response – if not properly accounted for – may thwart attempts to reliably estimate $DFLE$ and trends therein. This underscores the need for published estimates of $DFLE$ to be supplemented with a careful description of sample selection in the underlying survey and of the methods and data used to deal with this.

2. Recent studies on trends in $DFLE$ and attention paid to survey non-response

Increased longevity has generated a substantial interest in measuring health expectancies and in monitoring trends therein among the older population in developed countries.

[9] In Descriptive Appendix A.2, I discuss the (dis)similarities between Nord-Trøndelag and the entire country. The life expectancy figures shown there are fairly similar.

Saito, Robine, and Crimmins (2014) present an overview of the development in concepts and methods used to study and monitor changes in $DFLE$-like measures, and Ofstedal (2020) provides a comprehensive summary of available data sources for carrying out health-expectancy research.

Based on these methods and data, an enormous empirical literature has developed, estimating $DFLE$ and trends therein for many populations and time periods, and using different instruments to measure morbidity, disability, or functional limitation. Most of these studies differentiate between the two genders, and some also differentiate along other dimensions, such as race and educational attainment. Christensen et al. (2009) and Robine et al. (2020) provide extensive surveys of this empirical literature, while also giving the reader a bird's eye view of the trends. The first review, covering trends until about 2005, finds a generally increased prevalence of diseases in the elderly population, but also increasing evidence of falling disability prevalence. When the latter information is combined with mortality table information, however, the evidence is mixed. For example, in the cited study of Robine et al. (2005) that looks at $DFLE$ (at age 65 years) developments in 14 EU countries between 1995 and 2003, two countries are experiencing relative compression of disability for both genders (Belgium and Italy), two countries are experiencing stagnation (France and Spain), and two countries are witnessing a relative expansion of disability (Portugal and Netherlands).[10] However, the picture was less consistent for the remaining eight countries, with, e.g., German men (Swedish women) experiencing a relative compression of disability, whereas German women (Swedish men) experienced a relative expansion of morbidity. The survey by Robine et al. (2020) is at least a decade younger and therefore provides an even wider bird's eye view. They identify for the majority of countries (United States, United Kingdom, Sweden, Denmark, Norway, France, Spain, China, Australia) a relative compression of disability, though for some (such as the United Kingdom) expected lifetime with disability has expanded in absolute terms. A relative expansion of disability is reported for Japan, Hong Kong, and Singapore. Overall, these surveys point to a trend with relative compression of disability, but at the same time indicate a large heterogeneity around this trend.

Most country studies make use of the Sullivan method and combine information from the country's mortality tables with prevalence data (about disability or FI) from population surveys. These surveys often provide weights for adjustments intended to improve representativeness if data are missing because of non-response by invited individuals. The weights are usually based on external information available from the sampling frame. Most often such information is limited to age, gender, household type/size, and geographical location of residence, but, not seldom, only to age and gender. In Table A.1-1 of Descriptive Appendix A.1, I have listed 23 recent empirical studies of trends in health expectancy along with information about which survey the measures of health

[10] Absolute (relative) expansion or compression of morbidity/disability/FI means that the absolute expected time in this condition (this time relative to the expected lifetime) has increased or decreased, respectively.

status are obtained from, response rates in the underlying survey, and whether and how potential health-related unit non-response is addressed and dealt with.[11] Response rates vary considerably across surveys. Some studies provide information about response rates and use weights in the estimation of health status. Some inform that observations are weighted by population while referring to other sources for information about response rates, and some do not address this issue at all. Storeng et al. (2018, 2022), also using data from the HUNT Study, point at the declining response rates in the 'Discussion' section, but do not adjust for this in their analyses.

None of these studies address endogenous sample selection.

3. The data

My aim is to show how health-related survey non-responses impacts on the gender-specific age profiles for $FILE$ and its complement, $DFLE$. For this purpose, I make use of the abridged life tables for Norway (www.mortality.org), and data from the HUNT Study about the prevalence of FI in the Nord-Trøndelag county.

3.1 The HUNT Study

The HUNT Study[12] is a longitudinal health survey of the adult population of the former Nord-Trøndelag county (NT for short) in mid-Norway – now the northern part of Trøndelag county (see Descriptive Appendix A.2). The HUNT Study invited all NT residents aged 20 years and older for a medical check between January 1984 and April 1986 (the HUNT1 Survey: 86,404 residents), between August 1995 and June 1997 (the HUNT2 Survey: 93,897 residents) and between October 2006 and June 2008 (the HUNT3 Survey: 93,846 residents). In the sequel, I refer to these waves as HUNT1, HUNT2, and HUNT3, respectively. Each survey is described in detail by Holmen et al. (1990), Holmen et al. (2003), and Krokstad et al. (2013), respectively.

In every wave, all residents received an invitation by mail to meet at a field station on a particular date and time for a short health check. The invitation included the baseline questionnaire that people were asked to complete in advance and bring along. At the health check, participants underwent anthropometric measurements and provided blood and urine samples; they were also given a second questionnaire, which they were asked to return

[11] This list, which is not meant to be exhaustive, focuses on contributions published since 2000. It is based on studies reviewed by Robine et al. (2020) and supplemented by three Norwegian studies.

[12] The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Science, Norwegian University of Science and Technology NTNU), Trøndelag County Council, the Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. See https://www.ntnu.edu/hunt.

by mail. Participants gave their consent for their data to be linked with register data for research purposes. In this study, the survey data were linked to register data on age, gender, marital status, pension-entitling income, total income, and educational attainment.[13,14] The HUNT Study is a multi-purpose longitudinal dataset providing the empirical basis for many studies in (social) medicine and (public) health.[15]

The outcome variable, FI, indicates whether a respondent has a functional impairment ($FI = 1$) or not ($FI = 0$). It is based on the answer to the question "Do you suffer from any long-term illness or injury of a physical or psychological nature that impairs your functioning in your everyday life?" and that to the follow-up question about the severity of the impairment.[16] Figure 1 displays the gender-specific age profiles for FI for HUNT1 and HUNT3, based on the raw sample means, by solid lines.[17] All profiles are predominantly increasing, as expected, although some show a small dip around age groups 65 to 69/70 to 74 years.

[13] The HUNT Study complies with the principles of the Helsinki Declaration. All participants signed an informed consent to have their data merged with information from administrative registers. The current research project has been approved by the Committee for Medical and Health Research Ethics, REK Vest. Datasets are anonymous to researchers. Permissions to use HUNT and registry data does not extend to data sharing. The HUNT databank has precise information on all data exported to different projects and are able to reproduce these on request. There are no restrictions regarding data export, given approval of applications to the HUNT Research Centre and to the Regional Committees for Medical and Health research (REK: https://rekportalen.no/). For more information see http://www.ntnu.edu/hunt/data. Statistics Norway has merged the HUNT dataset with register data and owns the scrambling key (https://www.ssb.no/en/data-til-forskning/utlan-av-data-til-forskere).

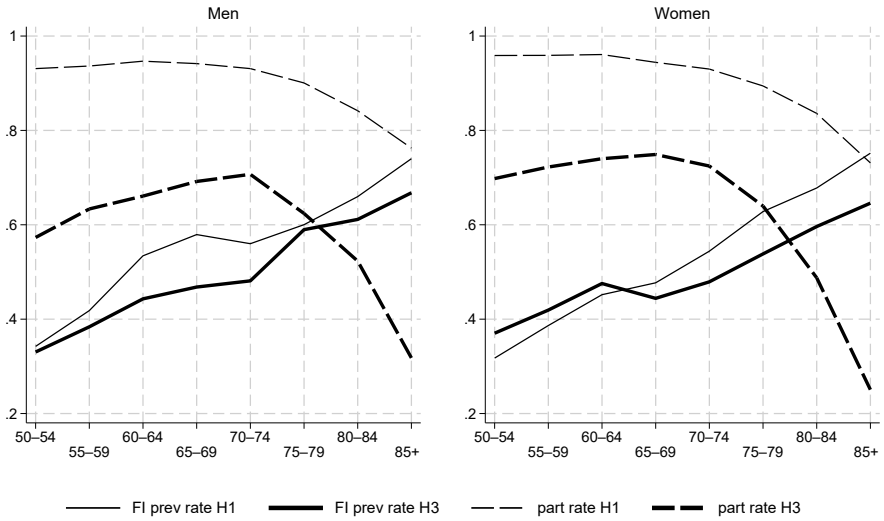
[14] With the exception of marital status, the registry data are also forwardly linked in the sense that for a resident who only participated in 1985 (HUNT1), registry data for 1996 and 2007 are available as well.

[15] At the time of writing, almost 1,500 research projects (concluded or ongoing) make use of the dataset and 826 articles using the HUNT data are listed in PubMed (https://pubmed.ncbi.nlm.nih.gov/?term=(trøndelag)+AND+(norway)+AND+(hunt)&sort=date).

[16] Details about the coding of the FI variable are provided in Descriptive Appendix A.3.

[17] In this and subsequent figures, the curves connect the five-year age group averages.

Figure 1: Age profiles for the prevalence rates for FI and the participation rate during HUNT1 and HUNT3



Note: The FI prevalence rate is based on the raw sample average. The participation rate is the number of answers to the FI indicator divided by the number of invitees.

At first glance, Figure 1 suggests that the prevalence of FI has fallen over time, especially for the higher-age groups. However, an alternative explanation is that a preponderance of functionally impaired residents has dropped out between HUNT1 and HUNT3, thereby giving the impression that the underlying population has become fitter. I have therefore supplemented the FI curves with the age profiles for the participation rates (long dashed lines). These are $\cap$-shaped; they tend to be somewhat higher (lower) for women than for men in age groups below 65 years (above 85 years); and they have declined over the 22-year period. These observations correspond to the stylized facts on participation in European health surveys (Mindell, Giampaoli, and Goesswald 2015) and lend credence to the alternative explanation.

I now discuss the construction of the working sample in more detail and pay particular attention to the attrition process.

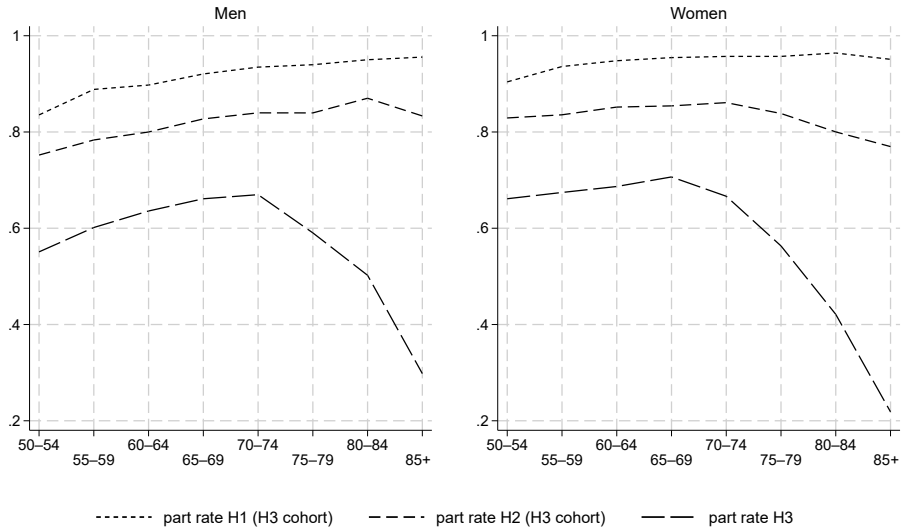
3.2 The working sample

The working sample consists of residents who participated in HUNT1 and either left the survey or participated in one or two consecutive surveys. Thus, I discard HUNT1 participants who did not participate in HUNT2 but returned to HUNT3 ('temporary attritors': 5.5%), as well as HUNT1 residents who joined only for HUNT2 and/or HUNT3 ('late joiners': 3.9%). The high participation rates for HUNT1 speak in favour of that sample being representative of the population of NT.[18] The maintained hypothesis is therefore that the prevalence of FI during HUNT1 can be estimated consistently based on the sample means in HUNT1.

However, the same cannot be assumed for HUNT3. There are two reasons why participants in HUNT1 did not participate in HUNT3. The first reason is attrition because of death or emigration from NT county.[19] This form of attrition is not worrisome as I am interested in measuring the prevalence of FI among the living residents. Because the dataset includes an indicator for being invited, I can condition on having survived from HUNT1 to HUNT3. I call the group of HUNT1 invitees who survived until HUNT3 the 'HUNT3-survivors.' The second reason is attrition of HUNT3-survivors between HUNT1 and HUNT3. This attrition is illustrated in Figure 2. This figure conditions on residents in NT who survived until HUNT3 and belonged to age groups 50 to 54, 55 to 59,..., 85+ years during HUNT3; thus it traces the same age cohorts over time. The upper line displays the participation rate of these cohorts during HUNT1. The middle line does the same for HUNT2, but now also conditioned on participation during HUNT1, whereas the lower line shows participation rates during HUNT3, conditioned on having participated during HUNT1 and HUNT2. Figure 2 shows clearly how the HUNT surveys struggle with falling participation of residents within the same age cohort. For example, of the 3,434 men who belonged to age group 50 to 54 years during HUNT3, 2,869 or 83.5% participated in HUNT1; of these 2,869 men, 75.2% participated in HUNT2 and only 55.1% also participated in HUNT3. Although Figure 2 clearly indicates that attrition is age-related, conditioning on age may not be sufficient to correct for non-response. Indeed, Langhammer et al. (2012) study participation in HUNT3 by linking HUNT3 data to follow-up questionnaires to non-participants, medical record data, and register data on socioeconomic characteristics. They report that non-participants had lower socioeconomic status, higher mortality, and higher prevalence of cardiovascular diseases, diabetes, and psychiatric disorders. The challenge is then to correct for the potential non-representativeness of the HUNT3-survivors' sample in HUNT3 because of health-related non-responses. I address this challenge in Section 4.

[18] A detailed follow-up study by Holmen et al. (1989) shows that attendance rates were lower among men, older people, and unmarried people, and that among the elderly, the non-attendants displayed higher mortality and morbidity rates than attendants. Although I will use HUNT1 as the benchmark sample, this should be kept in mind.
[19] Attrition refers to non-response in longitudinal surveys after the initial survey round or wave. Attrition is monotone when a non-responder never returns in later waves to the survey or non-monotone when a non-responder rejoins the survey in a later wave.

Figure 2: Age profiles for participation rates during HUNT1, HUNT2, and HUNT3, conditional on surviving until HUNT3 and on participating in the previous wave



Note: Participation rates for HUNT3-survivors and per age group reached under HUNT3. For HUNT2 and HUNT3, the rates are conditional on participating in the previous waves. HUNT3 age groups on horizontal axes.

Table 1 shows the working sample. I am interested in mapping the age profile for the prevalence rate of FI for age groups 50 to 54, 55 to 59,...,85+ years in HUNT1 and HUNT3. Row 1 describes the sample sizes for the 50+ years age subset of the population in HUNT1. For example, 20,831 female residents were invited, of whom 19,111 participated and answered the FI question. This gives a participation rate of 91.7%, suggesting a reliable sample for the purpose of measuring the FI prevalence rates during HUNT1. By way of illustration, the sample prevalence rate of FI for the age group 50 to 54 years is around one-third (column 4).

Table 1: Samples for measuring FI prevalence in HUNT1 (upper panel) and HUNT3 (lower panel)

wave	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					HUNT1 residents aged 50+ years during HUNT1						
	Invitees	Participants	Partic. rate	$\widehat{E(FI)}$ [a] $50-54$							
(1) HUNT1	19054*M*	17473*M*	91.7%*M*	0.342*M*							
	20831*W*	19111*W*	91.7%*W*	0.317*W*							
					HUNT3 residents aged 50+ years during HUNT3						
	Invitees	Participants	Partic. rate								
(2) HUNT1	19679*M*	17754*M*	90.2%*M*								
	21576*W*	20360*W*	94.4%*W*								
		↙ [b]									
	Invitees	Participants	Partic. rate					Permanent attritors	Sum (2) + (8)		(Temporary attritors) [c]
(3) HUNT2	17740*M*	14329*M*	80.8%*M*					2309*M*	16643*M*		(1097*M*
	20116*W*	16753*W*	83.3%*W*					2166*W*	18927*W*		(1189*W*)
		↙ [b]							↘ [b]		
	Invitees	Participants	Partic. rate	Partic.rate wrt (row2,col2)	$\widehat{E(FI)}$ [d] $50-54$	Non-participants	$\widehat{E(FI)}$ [e] $50-54$	Permanent attritors	$\widehat{E(FI)}$ [f] $50-54$	Sum (1) + (8)	$\widehat{E(FI)}$ [g] $50-54$
(4) HUNT3	14318*M*	10477*M*	73.2%*M*	59.1%*M*	0.330*M*	3841*M*	0.375*M*	2309*M*	0.417*M*	16627*M*	0.383*M*
	16735*W*	12025*W*	71.9%*W*	59.9%*W*	0.370*W*	4710*W*	0.433*W*	2166*W*	0.515*W*	18901*W*	0.442*W*

Note: [a] Raw sample average, that is, using estimator (1) over the participants, for $A = 50-54$. [b] Small differences between no. of participants in row 2 [3] and number of invitees in row 3 [4] are due to some of the explanatory variables being missing. [c] Temporary attritors are discarded (cf assumption of monotone attrition). [d] Raw sample average, that is, using estimator (1) over the participating survivors, for $A = 50-54$. [e] The average of (10) over the invitees (excl. those permanently attriting since HUNT2), for $A = 50-54$. [f] The average of (4.2) over the permanent attritors since HUNT2 for $A = 50-54$. [g] Using estimator (13) for $A = 50-54$.

Row 2 in Table 1 counts the number of invitees and participants in HUNT1 that survive til HUNT3 (HUNT3-Survivors) with 50+ years age during HUNT3. Because HUNT3 takes place about 22 years after HUNT1, the 20,360 participating women will also include women as young as age 26 years during HUNT1. Ideally, I would like to observe these 20,360 women (and 17,754 men) during HUNT3 and record their answer to the FI question in that survey. However, because of attrition between HUNT1 and HUNT3, a substantial number of these individuals did not participate in HUNT3. My aim is to correct for this attrition and to provide more consistent estimates for the FI age profiles and the corresponding $FILE$ profiles than the raw HUNT3 sample means would allow for.

I now explain how I trace these 20,360 women (and 17,754 men) further in time. Row 3 shows that 16,573 of these women participated in HUNT2.[20] Of the difference (3,363 women), about two-thirds became permanent non-responders, whereas about one-third of them (1,189) did not respond but reappeared again in HUNT3. As is common in longitudinal analysis, I assume monotone attrition and discard these temporary non-responders in the remainder of the analysis. For the 2,166 women (2,309 men) that permanently attrited, I propose in Section 4.2 an estimator for their FI age profile in HUNT3.

Row 4 shows that of the 16,573 participating women in HUNT2, 12,026 also participated during HUNT3. Compared with the target number identified in column 2 of row 2, this means a participation rate of only 59.1% (59.0% for men), meaning that about two-fifths of HUNT1 participants were lost because of attrition. To give a taste of the results below, the proposed method of correcting for this attrition results in an estimate for the FI prevalence rate of women invited to HUNT3 in the age group 50 to 54 years of 0.433 (column 7), and an estimate for the permanently attriting women in the same age group of 0.515 (column 9). With reference to the target of 19,171 women (the 20,360 identified in row 2, column 2 minus the 1,189 discarded temporary attritors), the estimate for the prevalence rate is 0.442 (0.383 for men) (column 11). This number can be compared with the raw sample estimate obtained from the HUNT3 participating women in the age group 50 to 54 years given in row 4, column 5: 0.370 (0.330 for men). In Section 5, I show the confidence regions for these results.

This concludes the sample description. For each respondent, I observe the outcome variable FI, a set of health status variables (BMI category, self-assessed health, incidence of diabetes, myocardial infarction, heart attack, and haemorrhage/stroke), as well as a list of socio-demographic variables (civil status, age, educational attainment, pension-entitling income, and total income). The last four of these are also available for non-respondents. The list of variables with descriptive statistics is given in Table A.4-1 of Descriptive Appendix A.4. I now explain the method for correcting for non-response.

[20] The minor differences between participants in row 2 [3] and invitees in row 3 [4] are because some background characteristics are missing in the later waves.

4. The statistical model

The outcome variable of interest is FI_{it} – namely, whether individual i is functionally impaired ($FI_{it} = 1$) or not ($FI_{it} = 0$) at the time when wave t of the survey is run. To simplify notation, I drop the wave subscript t. I am interested in the prevalence rate of FI in the population for age group A, $E(FI|A)$, where A denotes a five-year age group ($A = 50-54, 55-59, ..., 85+$ years). This rate is given by

$$\frac{\sum_{i \in \mathcal{P}(A)} FI_i}{\#\mathcal{P}(A)},$$

where $\mathcal{P}(A)$ denotes the set of individuals in the population belonging to age group A, and $\#\mathcal{P}(A)$ its cardinality. However, although all citizens are invited to the survey, not all of them participate. In the remainder of this section, I will explain in an intuitive way (1) why selection into a health survey is likely to be based on both observable and unobservable characteristics of invitees and, if not taken seriously, will produce biased estimates of FI prevalence rates, (2) how this selection can be corrected using a Heckman (1976) modelling approach, and (3) how, for a multi-purpose survey like HUNT, reliable estimation of this model can be conducted by letting the strength of selection on observables restrict the strength of selection on unobservables (Altonji, Elder, and Taber 2005). Emphasis will be on intuition, while all details and technicalities are spelled out in the technical appendix. Readers who directly want to learn about the consequences for age profiles of FI prevalence and $FILE/DFLE$ can jump to Section 5.

4.1 Selection on observables and unobservables

Let R_i be an indicator denoting whether individual i participates in the survey ($R_i = 1$) or not ($R_i = 0$). The prevalence rate among participants is then $E(FI|A, R = 1)$, which can be estimated by the sample analogue

$$\frac{\sum_{i \in \mathcal{S}(A)} FI_i}{\#\mathcal{S}(A)}, \tag{1}$$

where $\mathcal{S}(A)$ indicates the set of participants belonging to age group A. This estimator is available but not necessarily unbiased for $E(FI|A)$. To see this, observe that $E(FI|A)$ is a weighted average of $E(FI|A, R = 1)$ and $E(FI|A, R = 0)$, with weights $\Pr(R = 1|A)$ and $\Pr(R = 0|A)$, respectively. We know these weights (we can count how many invitees

in each age group participate or not), and we can estimate $E(FI|A, R = 1)$ by (1), but are agnostic about the expected prevalence of FI among non-participants, $E(FI|A, R = 0)$.

I now introduce three different assumptions, in increasing degree of weakness, to solve this problem. The first assumption, 'missingness completely at random' (MCAR), simply states that the mechanism driving non-response is purely random – independent of either the outcome variable FI or of any other variables in the dataset. This assumption could be defended if, say, the only reason for an invitee not showing up is that his/her invitation got lost in the mail. In that case $E(FI|A, R = 0) = E(FI|A, R = 1)$, and therefore (1) is an unbiased estimator of $E(FI|A)$. Obviously, MCAR is a very strong assumption unlikely to be satisfied.

A somewhat weaker assumption is that the probability of missingness still does not depend on FI, but may depend on a vector w of other variables in the dataset, such as socioeconomic variables and past health indicators. This weaker assumption is called 'missingness at random' (MAR) and means that $E(FI|A, w, R = 0) = E(FI|A, w, R = 1)$ for all vectors w. That is, once we condition on a sufficiently large set of characteristics, then respondents and non-respondents are on average alike in terms of FI. This case is also called selection on observables (the w vector).[21] The recipe to obtain an estimate for $E(FI|A)$ then goes as follows (see Technical Appendix B.1 for the details). First, MAR allows us to estimate the mean population prevalence for those with characteristics w by the mean prevalence rate of the respondents with these characteristics: $\widehat{E(FI|A, w)} \stackrel{\text{def}}{=} \frac{\sum_{i \in \mathcal{S}(A,w)} FI_i}{\#\mathcal{S}(A,w)}$. Second, a probit model for participation is estimated on the population of invitees, with the mean latent willingness to participate, R_i^*, a linear function of the characteristics vector w_i:

$$R_i = I(R_i^* = w_i'a + u_i > 0),\ u_i \sim N(0, 1), \tag{2}$$

where $I(\cdot)$ is the indicator function taking the value $1(0)$ if the expression in brackets is true (false). Invitee i then decides to participate if R_i^* exceeds zero. If $\widehat{a}$ denotes the maximum likelihood estimate of a, then the predicted participation probability for individual i with characteristics vector w_i is $p(w_i'\widehat{a}) = 1 - \Phi(-w_i'\widehat{a}) = \Phi(w_i'\widehat{a})$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Third, we construct a weighted average of the FI_i in the sample, where the weight for respondent i is proportional to the inverse of the participation probability, $\frac{1}{p(w_i'\widehat{a})}$. The intuition is that if individuals with characteristics vector w_i are underrepresented in the sample, then scaling up the FI prevalence rate of the few that are participating can make the sample representative again. This method is called

[21] The reason is that the equality $E(FI|A, w, R = 0) = E(FI|A, w, R = 1)$ can be inverted using Bayes' rule to give $E(R|A, w, FI = 0) = E(R|A, w, FI = 1)$: The probability of non-response (i.e., FI missing) does not depend on the value of FI once we control for the observables w. (Note that the expectation of a binary variable that can only take the values 0 or 1 coincides with the probability of taking the value 1.)

correcting for selection on observables by inverse-probability weighting. It results in the estimator

$$\sum_{i \in \mathcal{S}(A)} FI_i \frac{\frac{1}{p(w_i'\widehat{a})}}{\sum_{j \in \mathcal{S}(A)} \frac{1}{p(w_j'\widehat{a})}}. \tag{3}$$

However, in general, and in particular when the outcome variable of interest is a health variable, the probability of missingness will in addition also depend on the variable that we want to measure. Then, the data in the sample are no longer missing at random (called 'missing not at random' – MNAR – in the literature). In that case there are unobservable individual characteristics, which also determine FI, that may be correlated with the unobservables in the participation model (2) summarized in u. By stark example, imagine that all invitees in the health survey have an intention to participate but upon arrival at the survey facility learn that it is located on the second floor of a building without an elevator. Those with reduced mobility – likely to have $FI = 1$ – will then be inclined to return home, and their FI indicator will be missing in the dataset. More generally, suppose that the degree of FI of individual i, FI_i^*, is given by the linear model

$$FI_i^* = w_i'b + e_i, \tag{4}$$

where e_i summarises the effects of all unobservable characteristics, is also normally distributed with mean zero and variance σ_e^2, and has covariance with u_i equal to σ_{eu}. Using the properties of the bivariate normal distribution, the expected willingness to participate for an individual with observable characteristics w_i and unobservable characteristics e_i is

$$E(R_i^*|w_i, e_i) = w_i'b + E(u_i|e_i) = w_i'b + \frac{\sigma_{eu}}{\sigma_e^2} e_i,$$

showing clearly that selection into the sample depends on observables (w_i) and – to the extent that u_i and e_i are correlated ($\sigma_{eu} \neq 0$) – on unobservable FI components (e_i). To see the implications for the measurement of the prevalence rate of FI, we move our focus to (4). Suppose for a moment that we can observe FI_i^* (and not just $FI_i = I(FI_i^* > 0)$) when i participates. If we can obtain an unbiased estimate of $E(FI_i^*|A, w_i) = w_i'b$, we can proceed as in the MAR case. However, because we observe only FI_i for survey participants, the conditional mean of observed FI_i^* is not $w_i'b$ but

$$\begin{aligned} E(FI_i^*|A, w_i, R_i = 1) &= w_i'b + E(e_i|w_i, R_i = 1) = w_i'b + E(e_i|w_i, R_i^* > 0) \\ &= w_i'b + E(e_i|w_i, u_i > -w_i'a) = w_i'b + \sigma_{eu} \times E(u_i|u_i > -w_i'a), \end{aligned} \tag{5}$$

where the last equality follows from the joint normality of (e_i, u_i). Intuitively, we will observe individual i in the sample if u_i is large enough. However, if the correlation between u_i and e_i is negative, this individual's FI_i^* is likely to be smaller than $w_i'b$. In other words, participants have on average lower degrees of functional impairment.[22] The expectation $E(u_i|u_i > -w_i'a)$ equals $\frac{\phi(w_i'a)}{\Phi(w_i'a)}$, the ratio of the standard normal density function ($\phi(\cdot)$) to the standard normal cumulative distribution function, both evaluated at $w_i'a$ – also called the inverse Mills ratio. Expression (5) shows clearly that

$$E(FI_i^*|A, w_i, R_i = 1) = E(FI_i^*|A, w_i) + \sigma_{eu}\frac{\phi(w_i'a)}{\Phi(w_i'a)}, \tag{6}$$

and therefore that an estimate for $E(FI_i^*|A, w_i, R_i = 1)$ will under- (over-)estimate $E(FI_i^*|A, w_i)$ if $\sigma_{eu} < (>)0$. Only when the unobservables e and u are uncorrelated ($\sigma_{eu} = 0$ – the MAR case) will the sample mean of FI^* over those with characteristics w be an unbiased estimator of $E(FI^*|A, w)$.

Expression (6) identifies the heart of the problem of selection on unobservables, but also suggests a solution, introduced to the social sciences by Heckman (1976). Because (6) implies that

$$FI^*|_{w,R=1} = w'b + \sigma_{eu}\frac{\phi(w'a)}{\Phi(w'a)} + \zeta \text{ with } E(\zeta|w) = 0, \tag{7}$$

we can (1) replace $\frac{\phi(w'a)}{\Phi(w'a)}$ by a consistent estimate, $\frac{\phi(w'\widehat{a})}{\Phi(w'\widehat{a})}$; (2) estimate b consistently, $\widehat{b}$, by an ordinary least squares (OLS) regression of FI^* on w and $\frac{\phi(w'\widehat{a})}{\Phi(w,\widehat{a})}$; and (3) use $w'\widehat{b}$ as a consistent estimate for $E(FI^*|A, w)$. Next, we can proceed with steps 2 and 3 under MAR to obtain a consistent estimate for $E(FI^*|A)$.

There is one caveat to this solution, and it is related to step 2. The inverse Mills ratio is almost linear over a wide range in its argument, $w'a$ (see, e.g., Figure 1 in Puhani 2000: 57). This means that both the first and second right-hand side terms of (7) (with a replaced by $\widehat{a}$) are (almost) linear in w and therefore that OLS on (7) will suffer from collinearity: OLS does not manage to distinguish between the variation that comes from the w_i in the first and second term. This makes identification of b tenuous (Cameron and Trivedi 2005: 551). As I rely on a stable estimate for b to estimate $E(FI^*|A, w)$, the procedure needs to be rescued from this collinearity.

[22] One can also tell a positive correlation story. Individual i is fitter than the average NT citizen (e_i is negative). The day he/she is scheduled for the health check turns out to be a bright one with excellent cross-country skiing conditions. i decides to go skiing (u_i is sufficiently negative). Below, I will estimate the average correlation between e and u. This turns out to be negative.

The standard solution is to have at least one extra variable in the participation equation, or, stated differently, to have one variable in the vector w that is known to have a zero effect on FI^* (so that the corresponding coefficient in the vector b can be fixed at zero). If such a variable (coined 'instrument' in the literature) exists, then it will produce independent variation in the inverse Mills ratio and therefore guarantee the identification of the parameters of interest, b. In the earlier example, information about the floor level would qualify as an instrument. In that case, OLS on (6) will produce reliable estimates of b.

Unfortunately, most health surveys, including HUNT, do not include such an instrument. The solution that I therefore propose is to exploit the fact that the HUNT survey is a multi-purpose survey, designed with the aim of providing data on health characteristics but also socio-demographic information to support a wide variety of empirical research in social medicine and related disciplines. As explained in the introduction, the list of available variables to address a research question can then be regarded as a random selection from the entire list of variables necessary to answer that question. Such variables are then included in vector w, whereas others remain 'unobservable' and are summarized in the error term e. Altonji, Elder, and Taber (2002, 2005) then show that the strength of selection on unobservables (SoU) can be expected to be as large as that of selection on observables (SoO). Technical Appendix B.2 shows that 'SoU = SoO' amounts to $\frac{cov(u_i,e_i)}{var(e_i)}$ (i.e., $\frac{\sigma_{eu}}{\sigma_e^2}$) being equal to $\frac{cov(w_i'a,w_i'b)}{var(w_i'b)}$. This means that we can write the second right-hand side term of (7) as $\sigma_e^2 \frac{cov(w_i'\widehat{a},w_i'b)}{var(w_i'b)} \frac{\phi(w_i'\widehat{a})}{\Phi(w_i'\widehat{a})}$. Because $\frac{cov(w_i'\widehat{a},w_i'b)}{var(w_i'b)}$ is a non-linear function of w_i, it solves the collinearity problem and is sufficient to identify b (and σ_e^2) by OLS.[23]

This, in a nutshell, summarises how I deal with non-response in waves 2 and 3 of the HUNT survey. The formal approach that I take differs in two respects. First, because I do not observe FI^* but only the indicator FI, I use also a latent variable model for FI:

$$FI_i = I(FI_i^* = w_i'b + e_i > 0),$$

where the variance of e_i is now also normalised to one because only FI and not FI^* is observed.

Second, in the exposition so far, the vector of observables w was assumed to contain information on each invitee valid close to the time of the survey. However, the HUNT surveys are conducted about every 11th year. This means that information on many observable health variables will date back to 11 years earlier. The vector w can then be partitioned in a vector x with contemporaneous socio-economic information and previous wave health information, and a vector z with information about more recent health shocks

[23] When estimating, $cov(\cdot)$ and $var(\cdot)$ are replaced by the estimation sample covariance and variance, respectively. Since $\frac{cov(w_i'\widehat{a},w_i'b)}{var(w_i'b)}$ includes the coefficient vector b to be estimated, implementation will require an iterative process over OLS regressions or non-linear least squares estimation.

that occurred after the previous wave: $w' = (x', z')$. By definition, z is not observable for non-participants, and the effects of these unobserved variables on FI^* and R^* will together with the original error terms e_i and u_i result in a new pair of error terms ε_i and ν_i, respectively, with correlation ρ. We then obtain a reduced bivariate probit model

$$\begin{aligned} R_i &= I(R_i^* = x_i'\alpha + \nu_i > 0), \\ FI_i &= I(FI_i^* = x_i'\beta + \varepsilon_i > 0), \\ FI_i &\text{ observed if and only if } R_i = 1, \\ \begin{pmatrix}\nu_i \\ \varepsilon_i\end{pmatrix} &\sim N(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}). \end{aligned} \tag{8}$$

The fact that the new error terms ε_i and ν_i now consist of the original ones plus the effects of recent health shocks, and because it is reasonable to assume that such shocks matter more for selection than health conditions 11 years earlier, suggests that selection on unobservables will be stronger than selection on observables, and therefore that for the new reduced model $\frac{cov(x_i'\alpha, x_i'\beta)}{var(x_i'\beta)}$ constitutes a lower bound for ρ, that is, $\rho = \lambda \times \frac{cov(x_i'\alpha, x_i'\beta)}{var(x_i'\beta)}$ with $\lambda \geq 1$. I formally show this result in Technical Appendix B.2. In the empirical part of the article, the main results will be presented for $\lambda = 1$. These will be contrasted with results for $\lambda = 0$ (implying $\rho = 0$, the MAR case where selection into the sample only happens on observables) whereas Technical Appendix B.5 presents results for $\lambda = 1.2$. Technical Appendix B.3 explains how maximum likelihood estimation of (8) under the 'SoU=$\lambda\times$SoO' restriction can be done using an iterative procedure.

4.2 Implementation

I now explain how I implement the corrections for both forms of selection, while relegating the details to Technical Appendix B.4.

Step 1. I take the sample of HUNT1 participants as the benchmark and use estimator (1) on that sample of 19,111 (17,471) participating women (men) aged 50+ (years) in HUNT1 (row 1, column 2 in Table 1).

Step 2. I estimate model (8), under the restriction $\rho_2 = \frac{cov(x_{2i}'\alpha_2, x_{2i}'\beta_2)}{var(x_{2i}'\beta_2)}$, on the sample of 18,928 (16,644) invited and not temporarily attriting women (men) to HUNT2 (row 3, column 9 of Table 1). The relevant x_{2i}-vector consists of the variables listed in column 1 of Table 2. All these variables are observable for all invitees whether participating in HUNT2 or not because they either are recorded during HUNT1 or come from linked register data.

Table 2: List of explanatory variables when estimating model (8) for HUNT2 and HUNT3

vector x_2	vector x_3
age[a]	age[h]
civil status[b] in HUNT1	civil status in HUNT1
FI in HUNT1	*FI* in HUNT1
SAH-Good[c] in HUNT1	SAH-Good in HUNT1
BMI category[d] in HUNT1	BMI category in HUNT1
disease incidence[e] in HUNT1	disease incidence in HUNT1
	FI in HUNT2
educational attainment[f] in HUNT2	educational attainment in HUNT3
pension-entitling income in HUNT2	pension-entitling income in HUNT2
	pension-entitling income in HUNT3
permanent p-e income[g]	permanent p-e income
total income in HUNT2	total income in HUNT2
	total income in HUNT3

Notes: [a] Instead of entering age linearly, I use a more flexible restricted cubic age spline with seven knots determined by Harrell's percentiles (36, 42, 48, 55, 63, 73, 87) from age 35 years until the highest age in the data (the lowest age during HUNT2 for those belonging to the 50 to 54 year cohort during HUNT3 is 37 years). [b] Unmarried, divorced/separated, widowed (reference category: married). [c] Self-assessed health good or very good (reference category: poor or very poor). [d] BMI categories underweight, overweight, and obese (reference category: normal weight). [e] Indicators for having/having had diabetes, myocardial infarction, angina pectoris, stroke/cerebral haemorrhage. [f] Higher education, education missing (reference category: no higher education). [g] permanent p-e income is the average of p-e income over the three waves. [h] Instead of entering age linearly, I use a more flexible restricted cubic age spline with five knots determined by Harrell's percentiles (51, 58, 65, 73, 86) from age 50 years until the highest age in the data.

The estimation allows me to compute for every permanently attriting invitee the probability of having a FI conditional on not participating in HUNT2:

$$P(FI_{2i} = \widehat{1|x_{2i}}, R_{2i} = 0)\ i \in \mathcal{PA}_2, \tag{9}$$

where $\mathcal{PA}_2$ is the set of permanently attriting individuals since HUNT2. Probability (9) is useful for the next step.

Step 3. I estimate (8) under the SoU=SoO restriction $\rho_3 = \frac{cov(x'_{3i}\alpha_3, x'_{3i}\beta_3)}{var(x'_{3i}\beta_3)}$ on the sample of 16,735 (14,318) invited women (men) to HUNT3 (row 4, column 1 of Table 1). The relevant x_{3i}-vector consists of the variables listed in column 2 of Table 2. All these variables are observable for all invitees whether participating or not because they are either recorded during HUNT1 or HUNT2 (FI_{i2}) or come from linked register data. The estimation provides for all invitees to HUNT3 (i.e., all $i \in \mathcal{S}_2$), except those who are permanently attriting since HUNT2, the probability of being functionally impaired in

HUNT3 (which is now conditional on participation in HUNT2),

$$E(\widehat{FI_{3i}|x_{3i}})_{\text{H-AET}},\, i \in \mathcal{S}_2, \tag{10}$$

where the label H-AET refers to the Heckman-Altonji-Elder-Taber approach.

Moreover, I can borrow the estimated model to predict for each permanent attritor since HUNT2 the probability of being functionally impaired in HUNT3 given that individual had (did not have) FI in HUNT2:

$$P(FI_{3i} = \widehat{1|FI_{2i}} = 1, x_{3i}^{-})_{\text{H-AET}} \text{ and } P(FI_{3i} = \widehat{1|FI_{2i}} = 0, x_{3i}^{-})_{\text{H-AET}},\, i \in \mathcal{PA}_2, \tag{11}$$

where x_{3i}^{-} is the HUNT3 covariate vector excluding FI_{2i} – namely, $x'_{3i} = (FI_{2i}, x_{3i}^{-\prime})$. I then combine (9) and (11) to estimate for each such permanent attritor since HUNT2 the probability of having FI in HUNT3 as

$$\begin{aligned}
&E(\widehat{FI_{3i}|x_{3i}^{-}, \text{perm attr}})_{\text{H-AET}} = \\
&P(FI_{3i} = \widehat{1|FI_{2i}} = 1, x_{3i}^{-})_{\text{H-AET}} \times P(FI_{2i} = \widehat{1|x_{2i}}, R_{2i} = 0) + \\
&P(FI_{3i} = \widehat{1|FI_{2i}} = 0, x_{3i}^{-})_{\text{H-AET}} \times [1 - P(FI_{2i} = \widehat{1|x_{2i}}, R_{2i} = 0)],\ i \in \mathcal{PA}_2.
\end{aligned} \tag{12}$$

The estimator for the prevalence of FI to an HUNT3 survivor is therefore

$$\widehat{E(FI_{3i})}_{\text{H-AET}} = \frac{\sum_{i \in \mathcal{S}_2} E(\widehat{FI_{3i}|x_{3i}})_{\text{H-AET}} + \sum_{i \in \mathcal{PA}_2} E(\widehat{FI_{3i}|x_{3i}^{-}, \text{perm attr}})_{\text{H-AET}}}{\#\left(\mathcal{S}_2 \cup \mathcal{PA}_2\right)}. \tag{13}$$

Up to the temporary attritors in HUNT2, this estimator applies to all HUNT1 participants that are HUNT3 survivors; hence it corrects for attrition between HUNT1 and HUNT3. For example, for women in the age group 50 to 54 years, this estimate is 0.442 (given in row 4, column 11 of Table 2). It is the weighted average of 0.433 for participating women in HUNT2 (row 4, column 7) and of 0.515 for permanently attriting women since HUNT2 (row 4, column 9). Because of the relatively small number of permanent attritors since HUNT2, the estimates based on (13) are close to $\sum_{i \in \mathcal{S}_2} E(\widehat{FI_{3i}|x_{3i}})_{\text{H-AET}}/\#\mathcal{S}_2$.

5. Results for the FI age profiles

Given the assumption of the representativeness of HUNT1 participants, the wave 1 age profiles for FI can be estimated by (1). These were displayed in Figure 1 as thin solid lines.

To obtain the corresponding profiles for HUNT3, I proceed as explained above. The estimation results of (8) for HUNT2 and HUNT3 are shown in Tables 3 and 4, respectively, columns (1) and (2) for men and (3) and (4) for women. For ease of interpretation, these tables show the average partial effects with associated bootstrapped standard errors.[24] For example, having FI in the previous wave increases the probability of reporting being functionally impaired by around 24.5 to 26.4 percentage points but has a negligible effect on participation. Being divorced/separated in HUNT1 reduces participation in HUNT2 and HUNT3 for both men and women. Being obese in HUNT1 increases the probability of FI in HUNT2 and HUNT3 and reduces participation. A similar effect is found for being diagnosed with diabetes before HUNT1. The coefficients with the different income variables and with higher education confirm a positive gradient for health and for participation. The correlation coefficient ρ, measuring selection on unobservables and by construction equal to (the sample analogue of) $\frac{cov(x_i'\widehat{\alpha},x_i'\widehat{\beta})}{var(x_i'\widehat{\beta})}$ is estimated to be negative: On average, an unobserved shock that raises FI reduces the propensity to participate. Interestingly, the estimates for ρ suggest that selection was stronger for women than for men and more pronounced in HUNT3 than in HUNT2.

[24] The model was estimated by maximum likelihood using Stata/MP version 14.2. To obtain the bootstrapped standard errors, I estimated the model on 100 resamplings from the dataset and took the standard deviation of each coefficient. The Stata code is available upon request, as are the underlying estimated coefficients.

Table 3: Maximum likelihood estimations of H-AET model (8) on invitees to HUNT2 (aged 50+ years during HUNT3). Average partial effects

	(1)	(2)	(3)	(4)
	HUNT2 Men		HUNT2 Women	
	$\Pr(FI_2=1)$	$\Pr(R_2=1)$	$\Pr(FI_2=1)$	$\Pr(R_2=1)$
FI_1	0.264	0.014	0.245	0.001
	(0.013)	(0.008)	(0.011)	(0.006)
	[0.000]	[0.060]	[0.000]	[0.923]
unmarried$_1$	–0.024	–0.035	–0.008	–0.040
	(0.010)	(0.008)	(0.014)	(0.009)
	[0.021]	[0.000]	[0.593]	[0.000]
divorced$_1$	–0.005	–0.089	0.067	–0.051
	(0.022)	(0.020)	(0.018)	(0.014)
	[0.809]	[0.000]	[0.000]	[0.000]
widowed$_1$	–0.078	–0.034	–0.040	0.004
	(0.041)	(0.047)	(0.014)	(0.010)
	[0.056]	[0.467]	[0.005]	[0.665]
underweight$_1$	–0.052	–0.169	0.037	–0.028
	(0.078)	(0.072)	(0.027)	(0.023)
	[0.507]	[0.019]	[0.181]	[0.219]
overweight$_1$	0.009	–0.008	0.019	–0.018
	(0.007)	(0.005)	(0.008)	(0.006)
	[0.186]	[0.148]	[0.013]	[0.002]
obese$_1$	0.033	–0.046	0.037	–0.073
	(0.016)	(0.012)	(0.015)	(0.009)
	[0.038]	[0.000]	[0.012]	[0.000]
SAH Good$_1$	–0.110	–0.000	–0.139	–0.006
	(0.011)	(0.007)	(0.011)	(0.005)
	[0.000]	[0.980]	[0.000]	[0.297]
diabetes$_1$	–0.007	–0.076	0.033	–0.092
	(0.043)	(0.035)	(0.048)	(0.035)
	[0.866]	[0.031]	[0.488]	[0.009]
myoc. infarction$_1$	0.014	–0.018	–0.057	–0.057
	(0.043)	(0.034)	(0.066)	(0.068)
	[0.751]	[0.609]	[0.389]	[0.399]
angina pectoris$_1$	0.052	0.002	0.022	0.014
	(0.037)	(0.023)	(0.032)	(0.023)
	[0.165]	[0.916]	[0.507]	[0.529]
stroke$_1$	0.215	–0.036	0.032	–0.028
	(0.075)	(0.050)	(0.073)	(0.042)
	[0.004]	[0.469]	[0.664]	[0.504]
higher educ.$_2$	–0.010	0.019	–0.004	0.017
	(0.010)	(0.008)	(0.010)	(0.007)
	[0.285]	[0.015]	[0.654]	[0.014]
educ. missing$_2$	–0.016	–0.212	–0.056	–0.031
	(0.081)	(0.068)	(0.060)	(0.043)
	[0.844]	[0.002]	[0.354]	[0.470]

Table 3: (Continued)

	(1)	(2)	(3)	(4)
	HUNT2 Men		HUNT2 Women	
	$\Pr(FI_2=1)$	$\Pr(R_2=1)$	$\Pr(FI_2=1)$	$\Pr(R_2=1)$
p-e income$_2$	–0.005	–0.001	–0.013	0.001
	(0.002)	(0.001)	(0.001)	(0.001)
	[0.006]	[0.269]	[0.000]	[0.405]
p-e income (mean)	–0.026	0.014	–0.019	0.010
	(0.003)	(0.002)	(0.002)	(0.001)
	[0.000]	[0.000]	[0.000]	[0.000]
total income$_2$	–0.040	0.036	–0.036	–0.000
	(0.010)	(0.007)	(0.006)	(0.003)
	[0.000]	[0.000]	[0.000]	[0.995]
ρ	–0.079		–0.254	
	(0.021)		(0.021)	
	[0.000]		[0.000]	
Log likelihood	–13931.56		–15135.94	
Observations	16,643		18,927	
Uncensored obs	14,329		16,753	

*Note*s: Estimated on all HUNT3 survivors invited to HUNT2 and reaching age 50 years or higher during HUNT3; this means individuals in age interval [37–92] during HUNT2. Subscripts with the independent variables refer to HUNT waves (cf Table 2). A restricted cubic age spline with seven knots determined by Harrell's percentiles (36, 42, 48, 55, 63, 73, 87) is included among the regressors (not shown). SAH Good: takes the value 1 if "good" or "very good" and value zero if "not so good" or "poor." Stroke means stroke or cerebral haemorrhage. Reference category marital status: married. Reference category education: no higher education. p-e income: the inverse hyperbolic sine of the average pension-entitling income during the three survey years (HUNT2:1995–1997). p-e income (mean): the inverse hyperbolic sine of the mean of (average pension-entitling income during the three years HUNT survey years) over the HUNT surveys during which individual was present (but not necessarily participated). Total income: log of average real total income during the three survey years (HUNT2: 1995–1997). Standard errors obtained with bootstrapping (100 replications) in parentheses. *p*-values in square brackets.

The estimation results were fed into the estimator (13). Standard errors for this estimator were obtained by bootstrapping the entire procedure described in Section 4.2 (500 independent resamples from the dataset). The results are displayed by the solid curves ($E(FI)_{\texttt{H}-\texttt{AET}}$) in Figure 3, with associated 95% confidence area. These curves lie several percentage points above the long-dashed lines depicting the prevalence rates based on the raw HUNT3 sample mean of FI (i.e., using estimator (1)). The short-dashed lines in Figure 3 are constructed by relying only on inverse-probability weighting, which is based on the assumption that there is no selection on unobservable characteristics ($\lambda = \rho = 0$ – the MAR assumption).[25] It is clear that the correction stemming from the selection on observables alone is of minor order. Such a conclusion is empirically not uncommon –

[25] This estimator corresponds to (3) now with $p_i = \Phi(x'_{3i}\widehat{\alpha}_3) \times \Phi(x'_{2i}\widehat{\alpha}_2)$ – namely, the compound probability of participating both in HUNT2 and in HUNT3.

Table 4: Maximum likelihood estimations of H-AET model (8) on invitees to HUNT3 (aged 50+ years during HUNT3). Average partial effects

	(1)	(2)	(3)	(4)
	HUNT3 Men		HUNT3 Women	
	$\Pr(FI_3=1)$	$\Pr(R_3=1)$	$\Pr(FI_3=1)$	$\Pr(R_3=1)$
FI_2	0.261	0.006	0.260	–0.011
	(0.010)	(0.009)	(0.009)	(0.007)
	[0.000]	[0.525]	[0.000]	[0.113]
FI_1	0.133	0.007	0.119	–0.018
	(0.014)	(0.010)	(0.013)	(0.009)
	[0.000]	[0.499]	[0.000]	[0.053]
unmarried$_1$	–0.009	–0.052	0.003	–0.046
	(0.013)	(0.011)	(0.018)	(0.013)
	[0.501]	[0.000]	[0.860]	[0.000]
divorced$_1$	0.006	–0.116	0.087	–0.055
	(0.027)	(0.026)	(0.021)	(0.020)
	[0.834]	[0.000]	[0.000]	[0.005]
widowed$_1$	–0.002	–0.055	–0.034	–0.006
	(0.069)	(0.049)	(0.022)	(0.017)
	[0.977]	[0.264]	[0.130]	[0.726]
underweight$_1$	–0.078	0.027	0.054	–0.046
	(0.105)	(0.085)	(0.031)	(0.028)
	[0.458]	[0.747]	[0.083]	[0.097]
overweight$_1$	0.036	–0.017	0.037	–0.033
	(0.009)	(0.007)	(0.009)	(0.008)
	[0.000]	[0.019]	[0.000]	[0.000]
obese$_1$	0.086	–0.075	0.097	–0.090
	(0.019)	(0.017)	(0.015)	(0.013)
	[0.000]	[0.000]	[0.000]	[0.000]
SAH Good$_1$	–0.080	0.015	–0.077	–0.008
	(0.015)	(0.011)	(0.013)	(0.010)
	[0.000]	[0.201]	[0.000]	[0.427]
diabetes$_1$	0.130	–0.040	0.203	–0.022
	(0.070)	(0.048)	(0.055)	(0.042)
	[0.063]	[0.405]	[0.000]	[0.598]
myoc. infarction.$_1$	0.135	–0.040	0.040	–0.158
	(0.071)	(0.049)	(0.874)	(0.118)
	[0.059]	[0.412]	[0.963]	[0.182]
angina pectoris$_1$	–0.060	0.013	0.031	0.041
	(0.048)	(0.031)	(0.057)	(0.032)
	[0.215]	[0.685]	[0.581]	[0.204]
stroke$_1$	–0.025	–0.149	–0.014	–0.148
	(0.111)	(0.069)	(0.100)	(0.061)
	[0.823]	[0.031]	[0.892]	[0.015]
higher educ.$_3$	–0.036	0.043	–0.039	0.046
	(0.013)	(0.010)	(0.012)	(0.009)
	[0.005]	[0.000]	[0.001]	[0.000]

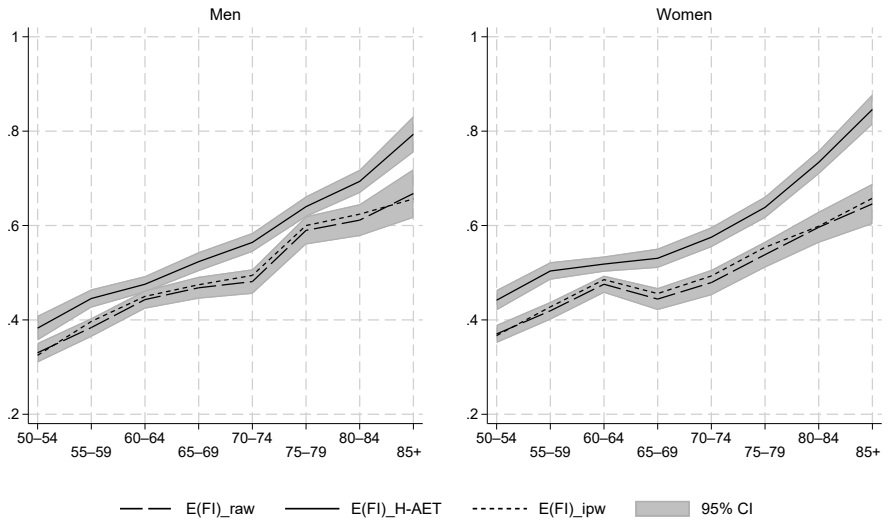
Table 4: (Continued)

	(1)	(2)	(3)	(4)
	HUNT3 Men		HUNT3 Women	
	$\Pr(FI_3 = 1)$	$\Pr(R_3 = 1)$	$\Pr(FI_3 = 1)$	$\Pr(R_3 = 1)$
educ. missing$_3$	–0.136	–0.156	–0.020	–0.040
	(0.331)	(0.123)	(0.077)	(0.054)
	[0.680]	[0.203]	[0.798]	[0.454]
p-e income$_3$	–0.007	–0.003	–0.016	0.001
	(0.004)	(0.002)	(0.001)	(0.001)
	[0.046]	[0.125]	[0.000]	[0.204]
p-e income (mean)	–0.018	0.024	–0.000	0.012
	(0.010)	(0.006)	(0.003)	(0.002)
	[0.068]	[0.000]	[0.928]	[0.000]
total income$_3$	–0.087	0.031	–0.001	0.008
	(0.016)	(0.012)	(0.007)	(0.005)
	[0.000]	[0.011]	[0.912]	[0.125]
p-e income$_2$	–0.007	–0.002	–0.006	0.004
	(0.004)	(0.002)	(0.002)	(0.001)
	[0.083]	[0.475]	[0.001]	[0.003]
total income$_2$	0.014	0.036	0.020	–0.002
	(0.014)	(0.011)	(0.006)	(0.005)
	[0.290]	[0.001]	[0.001]	[0.759]
ρ	–0.350		–0.536	
	(0.026)		(0.035)	
	[0.000]		[0.000]	
Log likelihood	–13907.72		–15735.04	
Observations	14,318		16,735	
Uncensored obs	10,477		12,025	

Notes: Estimated on all HUNT3 survivors invited to HUNT3 and reaching age 50 years or higher during HUNT3; the age interval is [50, 101]. Subscripts with independent variables refer to HUNT waves (cf Table 2). A restricted cubic age spline with five knots determined by Harrell's percentiles (51, 58, 65, 73, 86) is included among the regressors (coefficients not shown). SAH Good: takes the value 1 if "good" or "very good" and value zero if "not so good" or "poor." Stroke means stroke or cerebral haemorrhage. Reference category marital status: married. Reference category education: no higher education. p-e income: the inverse hyperbolic sine of the average pension acquiring income during the three survey years (HUNT2:1995–1997, HUNT3:2006–2008). p-e income (mean): the inverse hyperbolic sine of the mean of average pension acquiring income during the three years HUNT survey years over the HUNT surveys during which individual was present (but not necessarily participated). Total income: log of average real total income during the three survey years (HUNT2:1995–1997, HUNT3:2006–2008). Standard errors obtained with bootstrapping (100 replications) in parentheses. *p*-values in square brackets.

see, for example, the recent study by Nilsson et al. (2020).[26] In the next section, I discuss the consequences for measuring life expectancy with and without FI and the evolution in these measures during the time period 1985–2007.

Figure 3: Age profiles for the prevalence rate of FI during HUNT3



Notes: $E(FI)_{\texttt{raw}}$ uses estimator (1). $E(FI)_{\texttt{ipw}}$ uses estimator (3). $E(FI)_{\texttt{H-AET}}$ uses estimator (13). The 95%-confidence areas belonging to $E(FILE)_{\texttt{raw}}$ and $E(FILE)_{\texttt{H-AET}}$ are computed on 500 replications of the data set.

6. Trends in disability-free life expectancy

$DFLE$ is defined as the number of years that a person is expected to continue to live without FI. The Sullivan (1971) method applies the prevalence rates of FI for age interval $[a, a+4)$ to the person-years lived during that interval. Dividing the total number of person-years lived with FI after age a by the survivors until age a gives the functionally

[26] In their study on the dynamics of self-assessed health in the BHPS and the ECHP, Jones, Koolman, and Rice (2006) account for health-related non-response by inverse-probability weighting. They find that the differences between the coefficient estimates based on *ip* weighting and the regular estimates are not statistically different from zero and conclude that although there is clear evidence of health-related non-response, "on the whole it does not distort the magnitudes of the estimated dynamics of SAH and the relationship between socio-economic status [variables] and SAH" (p. 567).

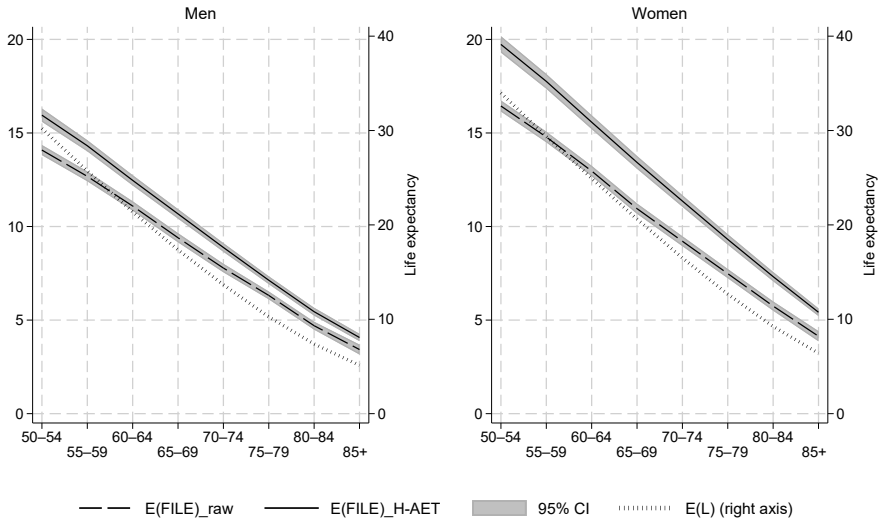
impaired life expectancy ($FILE$) for age group $[a, a+4)$.[27] Subtracting this from the total expected lifetime (LE) results in $DFLE$ for that age group.

I compute $FILE$ and $DFLE$ for men and women for five-year age groups starting at age 50 years with 85+ as the final open-end interval.[28] As explained in the introduction, I substitute the life tables for Norway for the ones for NT (cf Descriptive Appendix A.2). Because I assume that HUNT1 is representative, I have only a single estimator for $FILE$ in HUNT1, based on the raw sample prevalence rates of FI, $E(FI)_{\text{raw}}$. For HUNT3, I have, in addition, the estimator when only correcting for selection on observables, based on $E(FI)_{\text{ipw}}$, and when also controlling for selection on unobservables, based on $E(FI)_{\text{H-AET}}$. However, the results in the previous section showed that only correcting for selection on observables has very little impact on the age profiles for FI, and calculations (available upon request) show that this carries over to the age profiles for $FILE$ and $DFLE$. I therefore focus on comparing the age profiles for $FILE_{\text{raw}}$ and $FILE_{\text{H-AET}}$. These are displayed in Figure 4, along with 95% confidence bands based on the bootstrap procedure mentioned above. For women in age group 50 to 54 years, $FILE_{\text{raw}}$ is estimated at 16.4 years whereas $FILE_{\text{H-AET}}$ is 3.3 years higher. For men in age group 50 to 54 years, taking into account endogenous selection, adds almost two years with respect to $FILE_{\text{raw}}$. Figure 5 shows that correcting for selection affects $DFLE$ adversely. In Technical Appendix B.5, I present the results of a sensitivity analysis, assuming that for HUNT3 the strength of selection on unobservables is 120% (rather than 100%) that of selection on observables. The result is a further drop in $DFLE$ at age 50 to 54 years in HUNT3 of 0.4 years for men and 0.5 years for women.

[27] The Sullivan method is explained in detail by Jagger, Van Oyen, and Robine (2014). See also Imai and Soneji (2007) for a thorough discussion of its statistical properties.

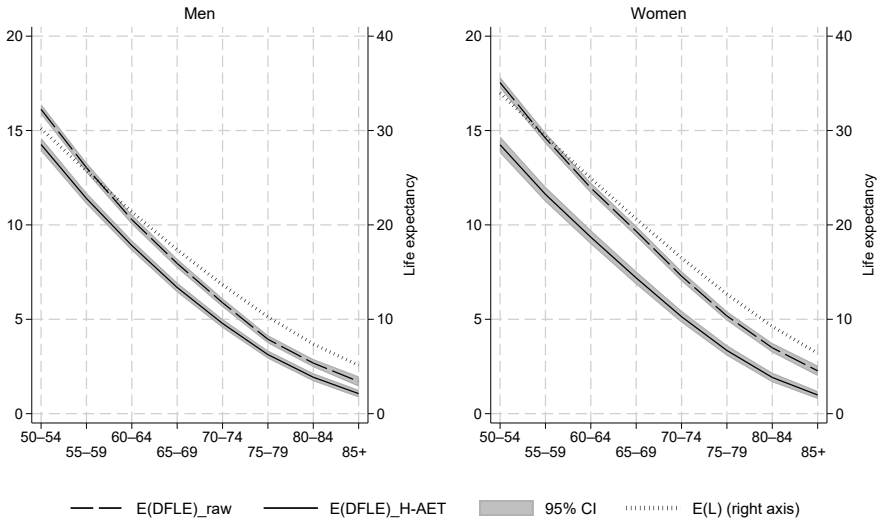
[28] Storeng et al. (2018) present estimates of changes in $DFLE$ based on changes over time in crude proportions with FI. My estimates of the FI prevalence rates are not directly comparable because (1) the outcome variable FI is defined slightly differently (see Section 3 and Descriptive Appendix A.3), (2) they report $DFLE$ at age 30 years, and (3) they calculate LE using mortality data for participants in the survey rather than using life tables for the population. Moreover, they do not correct for selection.

Figure 4: Age profiles for $FILE$ during HUNT3



Notes: The Sullivan method was employed. $E(FILE)_{\texttt{raw}}$ is based on the age profile given by $E(FI)_{\texttt{raw}}$. $E(FILE)_{\texttt{H}-\texttt{AET}}$ is based on the age profiles for $E(FI)_{\texttt{H}-\texttt{AET}}$. The 95% confidence areas belonging to $E(FILE)_{\texttt{raw}}$ and $E(FILE)_{\texttt{H}-\texttt{AET}}$ are computed on 500 replications of the data set. $E(L)$ as reported on www.mortality.org.

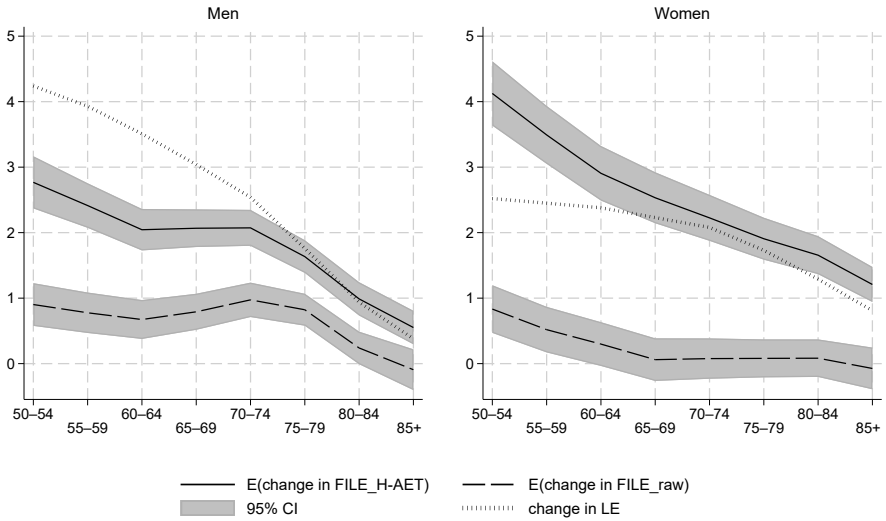
Figure 5: Age profiles for $DFLE$ during HUNT3



Notes: The Sullivan method was employed. $E(DFLE)_{\texttt{raw}} = LE - E(FILE)_{\texttt{raw}}$, $E(DFLE)_{\texttt{H-AET}} = LE - E(FILE)_{\texttt{H-AET}}$. The 95% confidence areas belonging to $E(DFLE)_{\texttt{raw}}$ and $E(DFLE)_{\texttt{H-AET}}$ are computed on 500 replications of the dataset. $E(L)$ as reported on www.mortality.org.

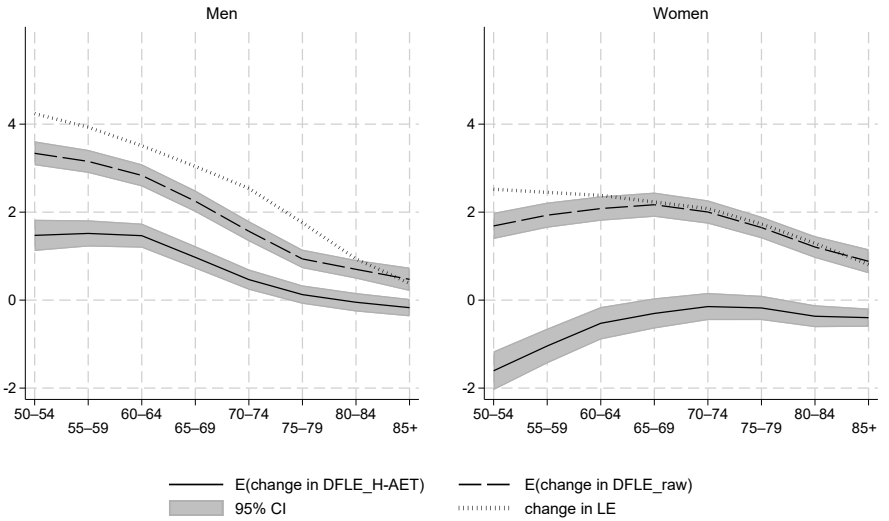
During the period 1985–2007, Norway witnessed a substantial increase in life expectancy. It increased from 31.5 to 34.0 (+2.5) years at age 50 to 54 years for women and from 26.0 to 30.2 (+4.2) years for men. The dotted lines in Figure 6 show the age profiles for this change in LE. When I rely on the raw sample FI prevalence rates, $FILE$ has increased with less than a year for the 50 to 54 year old group (long dashed lines). The mirror image is an increase in $DFLE$ that almost matches the increase in LE – see Figure 7. However, controlling for health-related selection, the picture is less bright: For men, the 4.2 years increase in LE comes with an estimated increase in $FILE$ of 2.8 years, leaving them with an estimated increase in $DFLE$ of 1.4 years, whereas for women the estimated increase in $FILE$ more than outweighs the increase in LE, thus giving a reduction in estimated $DFLE$.

Figure 6: Age profiles for the change in $FILE$ and LE (1985–2007)



Notes: Based on the HUNT3 age profiles drawn in Figure 4 and the age profile $E(FILE)_{\texttt{raw}}$ for HUNT1. The 95% confidence areas belonging to $E($change in $FILE)_{\texttt{raw}}$ and $E($ change in $FILE)_{\texttt{H}-\texttt{AET}}$ are computed using 500 resamples of the dataset. Change in $E(L)$ based on www.mortality.org.

Figure 7: Age profiles for the change in $DFLE$ and LE (1985–2007)



Notes: Based on the HUNT3 age profiles drawn in Figure 5 and the age profile for $E(DFLE)_{\texttt{raw}}$ for HUNT1. The 95% confidence areas belonging to $E($ change in $DFLE)_{\texttt{raw}}$ and $E($ change in $DFLE)_{\texttt{H}-\texttt{AET}}$ are computed using 500 resamples of the dataset. Change in $E(L)$ based on www.mortality.org.

7. Concluding remarks

The measurement of changes over time in health expectancy, or its complement, life expectancy with a disability or ill health, is challenging because the most important source of information about population health, health survey data, suffers from health-related unit non-response and falling response rates over time. This produces biased estimates of prevalence rates of disability or illness.

I make use of three waves of the Trøndelag Health Study (HUNT) to estimate changes in functional impairment in the population aged 50+ years over the 22-year period from 1985 to 2007. Because of exceptionally high response rates in the first wave, I start by assuming that initial non-response bias is negligible. For consecutive waves, I include a broad set of observed individual socio-demographic and health characteristics collected from the survey questionnaires, the health examinations, and from register data. I compare inverse-probability weighted prevalence rates of FI, which rest on the assumption that selection because of non-response is based on only observable characteristics, with bivariate probit estimates of FI where I also allow for selection on unobservable factors

that may be correlated with health. I restrict the model by assuming that selection on unobservables is as strong as selection on observables (Heckman 1976, 1979; Altonji, Elder, and Taber 2005).

The results confirm that several indicators of health limitations measured at younger ages significantly reduce the propensity to participate in the second and third waves of the HUNT Study. For both genders, obesity, diabetes, and earlier histories of stroke or cerebral haemorrhage predict non-response, whereas good self-assessed health predicts the opposite.

Subject to the restriction that selection on unobservables is as strong as on observables, I obtain negative estimates for the correlations between the error components in the equations for response and functional impairment, and find that these increase (in absolute value) from the second to the third HUNT wave. This results in age profiles for $DFLE$ that clearly lie below the same profiles when selection is assumed to take place only on observables or is assumed away altogether.

What is the most plausible assumption about the relative degree of selection? Without access to more data, this question cannot be answered, and my results should therefore be regarded as a sensitivity analysis exploring the consequences of raising the relative degree from zero to one. In general, the relative degree of selection will depend on the time span between health surveys, the effort that is undertaken to convince people to participate, and the format of the survey (e.g., whether or not participation requires physical attendance at a field station). These issues are rarely discussed in the context of $DFLE$ measurement and deserve attention in future research.

8. Acknowledgments

I am grateful to two anonymous reviewers for their constructive comments and suggestions. Along the way of this project, I have benefited from comments and collegial help from Liam Brunt, Egil Kjerstad, Lukas Laffers, Øivind Anti Nilsen, Gaute Torsvik, and seminar participants at the University of Malaga and NHH. Special thanks to Jonas Andersson for sharing his statistical expertise and to Astrid Hanssen Wang for long discussions on the subject, help with data management, and sharing her knowledge of the literature.

References

Altonji, J., Elder, T., and Taber, C. (2002). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Mimeo, Department of Economics, Northwestern University.

Altonji, J., Elder, T., and Taber, C. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113(1): 151–184. doi:10.1086/426036.

Cai, L. and Lubitz, J. (2007). Was there a compression of disability for older Americans from 1992 to 2003? *Demography* 44(3): 479–495. doi:10.1353/dem.2007.0022.

Cambois, E., Robine, J.-M., and Brouard, N. (1999). Life expectancies applied to specific statuses. A history of the indicators and the methods of calculation. *Population: An English Selection* 11: 7–34. doi:10.3917/popu.p1999.11n1.0034.

Cameron, C. and Trivedi, P. (2005). *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511811241.

Christensen, K., Doblhammer, G., Rau, R., and Vaupel, J.W. (2009). Ageing populations: The challenges ahead. *Lancet* 374(9696): 1196–1208. doi:10.1016/S0140-6736(09)61460-4.

Contoyannis, P., Jones, A., and Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics* 19(4): 473–503. doi:10.1002/jae.755.

Czajka, J.L. and Beyler, A. (2016). Declining response rates in federal surveys: Trends and implications. Washington, D.C.: Mathematica Policy Research.

De Carvalho Yokota, R.T. and Van Oyen, H. (2020). Operationalization of concepts of health and disability. In: Jagger, C., Crimmins, E., Saito, Y., De Carvalho Yokota, R., Van Oyen, H., and Robine, J.-M. (eds.). *International handbook of health expectancies*. Cham: Springer: 3–18. doi:10.1007/978-3-030-37668-05.

De Luca, G. and Perotti, V. (2011). Estimation of ordered response models with sample selection. *Stata Journal* 11(2): 213–239. doi:10.1177/1536867X1101100204.

European Commission (2018). The 2018 ageing report: Economic and budgetary projections for the EU member states (2016–2070). EC Institutional Paper 079. Brussels: European Commission.

Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3): 130–135. doi:10.1056/NEJM198007173030304.

Fries, J.F. (2003). Measuring and monitoring success in compressing morbidity. *Annals of Internal Medicine* 139(5, part 2): 455–459. doi:10.7326/0003-4819-139-5_Part_2-200309021-00015.

Greene, W. (2018). *Econometric analysis*. London: Pearson.

Gruenberg, E. (1977). The failures of success. *The Milbank Memorial Fund Quarterly. Health and Society* 55(1): 3–24. doi:10.2307/3349592.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator of such models. *Annals of Economic and Social Measurement* 5(4): 475–492.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1): 153–161. doi:10.2307/1912352.

Holmen, J., Forsèn, L., Skjerve, K., Gorseth, M., Midthjell, K., and Oseland, A. (1989). Møter – ikke møter? Helseundersøkelsen i Nord-Trøndelag 1984–86. Sammenliknende analyse av de som møtte og de som ikke møtte. [Attend – not attend? The Health Study in Nord-Trøndelag County 1984–85: Analysis of attendants and non attendants]. Rapport 5. Oslo: Statens institutt for folkehelse, avdeling for helsetjenesteforskning.

Holmen, J., Midthjell, K., Forsèn, L., Skjerve, K., Gorseth, M., and Oseland, A. (1990). A health survey in Nord-Trøndelag 1984–1986. Participation and comparison of attendants and non-attendants. *Journal of the Norwegian Medical Association (Tidsskrift for Den norske legeforening)* 110(15): 1973–1977.

Holmen, J., Midthjell, K., Krüger, Ø., Langhammer, A., Holmen, T.L., Bratberg, G.H., and Lund-Larsen, P.G. (2003). The Nord-Trøndelag Health Study 1995–97 (HUNT 2): Objectives, contents, methods and participation. *The Norwegian Journal of Epidemiology (Norsk Epidemiologi)* 13(1): 19–32. doi:10.5324/nje.v13i1.305.

Imai, K. and Soneji, S. (2007). On the estimation of disability-free life expectancy. *Journal of the American Statistical Association* 102(480): 1199–1211. doi:10.1198/016214507000000040.

Jagger, C., Crimmins, E.M., Saito, Y., De Carvalho Yokota, R.T., Van Oyen, H., and Robine, J.-M. (2020). *International handbook of health expectancies*. Cham: Springer. doi:10.1007/978-3-030-37668-0.

Jagger, C., Van Oyen, H., and Robine, J.-M. (2014). *Health expectancy calculations by the Sullivan method: A practical guide*. Newcastle upon Tyne: Newcastle University, 4th ed. http://www.eurohex.eu/pdf/Sullivan_guide_pre%20final_oct%202014.pdf.

Jones, A., Koolman, X., and Rice, N. (2006). Health-related non-response in the British Household Panel Survey and European Community Household Panel: Using inverse-

probability weighted estimators in non-linear models. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 169(3): 543–569. doi:10.1111/j.1467-985X.2006.00399.x.

Kramer, M. (1980). The rising pandemic of mental disorders and associated chronic diseases and disabilities. *Acta Psychiatrica Scandinavica* 62(S285): 382–397. doi:10.1111/j.1600-0447.1980.tb07714.x.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordeo, C., Lemay, M., Peytcheva, E., Groves, R.M., and Raghunathan, T.E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 173(2): 389–407. doi:10.1111/j.1467-985X.2009.00621.x.

Krokstad, S., Langhammer, A., Romundstad, P., Heggland, J., and Holmen, J. (2013). Cohort profile: The HUNT Study, Norway. *International Journal of Epidemiology* 42(4): 968–977. doi:10.1093/ije/dys095.

Lamb, V.L. and Siegel, J.S. (2004). Health demography. In: Siegel, J.S. and Swanson, D.A. (eds.). *The methods and materials of demography*. Amsterdam: Elsevier: 341–370.

Langhammer, A., Krokstad, S., Romundstad, P., Heggland, J., and Holmen, J. (2012). The hunt study: Participation is associated with survival and depends on socioeconomic status, diseases and symptoms. *BMC Medical Research Methodology* 12(143): 1–14. doi:10.1186/1471-2288-12-143.

Manton, K. (1982). Changing concepts of morbidity and mortality in the elderly population. *The Milbank Memorial Fund Quarterly. Health and Society* 60(2): 183–244. doi:10.2307/3349767.

Mathers, C.D. (2002). Health expectancies: An overview and critical appraisal. In: Murray, C., Salomon, J., Mathers, C.D., and Lopez, A. (eds.). *Summary measures of population health*. Geneva: WHO: 177–204.

Mathers, C.D., Sadana, R., Salomon, J., Murray, C., and Lopez, A. (2000). Estimates of DALE for 191 countries: Methods and results. WHO (Global Programme on Evidence for Health Policy Working Paper no. 16).

Matthews, F.E., Arthur, A., Barnes, L.E., Bond, J., Jagger, C., Robinson, L., and Brayne, C. (2013). A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: Results from the Cognitive Function and Ageing Study I and II. *The Lancet* 382(9902): 1405–1412. doi:10.1016/S0140-6736(13)61570-6.

Meyer, B., Mok, W.K.C., and Sullivan, J. (2015). Household surveys in crisis. *Journal of Economic Perspectives* 29(4): 199–226. doi:10.1257/jep.29.4.199.

Mindell, J., Giampaoli, S., and Goesswald, A. (2015). Sample selection, recruitment and participation rates in health examination surveys in Europe – Experience from seven national surveys. *BMC Medical Research Methodology* 15(78): 1–19. doi:10.1186/s12874-015-0072-4.

Murray, C. and Lopez, A. (eds.) (1996). *The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020.* Cambridge: Cambridge University Press.

Nilsson, A., Bonander, C., Stömberg, U., and Björk, J. (2020). Can the validity of a cohort be improved by reweighting based on register data? Evidence from the Swedish MDC Study. *BMC Public Health* 20(1918). doi:10.1186/s12889-020-10004-z.

Ofstedal, M.B. (2020). Data sources for health expectancy research. In: Jagger, C., Crimmins, E., Saito, Y., De Carvalho Yokota, R., Van Oyen, H., and Robine, J.-M. (eds.). *International handbook of health expectancies*. Cham: Springer: 69–86. doi:10.1007/978-3-030-37668-05.

Puhani, P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14(1): 53–68. doi:10.1111/1467-6419.00104.

Robine, J.-M. (2002). A new health expectancy classification system. In: Murray, C., Salomon, J., Mathers, C., and Lopez, A. (eds.). *Summary measures of population health*. Geneva: WHO: 205–211.

Robine, J.-M., Jagger, C., Crimmins, E.M., Saito, Y., and Van Oyen, H. (2020). Trends in health expectancies. In: Jagger, C., Crimmins, E.M., Saito, Y., De Carvalho Yokota, R.T., Van Oyen, H., and Robine, J.-M. (eds.). *International handbook of health expectancies*. Cham: Springer: 19–34. doi:10.1007/978-3-030-37668-0_2.

Robine, J.-M., Jagger, C., Van Oyen, H., Cambois, E., Romieu, I., Clavel, A., Barker, G., and Le Roy, S. (2005). Are we living longer, healthier lives in the EU? Disability-free life expectancy (DFLE) in EU countries from 1991 to 2003 based on the European Household Panel. EHEMU (Technical Report 2). Montpellier: EHEMU: 1–29.

Saito, Y., Robine, J.-M., and Crimmins, E.M. (2014). The methods and materials of health expectancy. *Statistical Journal of the IAOS* 30(3): 209–223. doi:10.3233/SJI-140847.

Sargent, T.A. (1976). Extensions to notes on macroeconomics. Minnesota: Federal Reserve Bank of Minneapolis (Working Paper 64).

Storeng, S.H., Krokstad, S., Westin, S., and Sund, E.R. (2018). Decennial trends and inequalities in healthy life expectancy: The HUNT Study, Norway. *Scandinavian Journal of Public Health* 46(1): 124–131. doi:10.1177/1403494817695911.

Storeng, S.H., Øverland, S., Skirbekk, V., Hopstock, L.A., Sund, E.R., Krokstad, S., and Strand, B.H. (2022). Trends in disability-free life expectancy (DFLE) from 1995 to 2017 in the older Norwegian population by sex and education: The HUNT Study. *Scandinavian Journal of Public Health* 50(5): 542–551. doi:10.1177/14034948211011796.

Sullivan, D.F. (1971). A single index of mortality and morbidity. *HSMHA Health Reports* 86(4): 347–354. doi:10.2307/4594169.

Van de Ven, W.P. and Van Praag, B.M. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics* 17(2): 229–252. doi:10.1016/0304-4076(81)90028-2.

A. Descriptive appendix

A.1 Literature survey

Table A.1-1 lists 22 recent empirical studies of trends in health expectancies along with information about which survey the measures of health status are obtained from, response rates in the underlying survey, and whether and how potential health-related unit non-response is addressed and dealt with. This list, which is not meant to be exhaustive, focusses on contributions published since 2000. It is based on studies reviewed by Robine et al. (2020) and supplemented with three Norwegian studies.

Table A.1-1: Analyses of *DFLE*/Healthy life expectancy. Sources of health information, response rates, and adjustment for non-response

Publication	Survey	Country/ Region	Age	Period	Response rates	Adjustments
Cai and Lubitz (2007)	Medicare Current Beneficiary Survey + mortality data	USA	65+, 85+	1992–2003	Declining from 87.2 to 78% at initial interview, vary between 90-95% at follow up[a]	No
Jeune and Brønnum-Hansen (2008)	The Danish Health Interview Survey	Denmark	65+	1987–2005	Decreasing from 76.5 to 64%	No
Crimmins et al. (2009)	The Longitudinal Studies of Ageing I and II	USA	70+	1984–2000	93.2% and 87.4% at baseline	Weights
Crimmins and Beltrán-Sánchez (2010)	The National Health Interview Survey	USA	20+, 65+	1998, 2006	1998: 90% household, 83.8% adult, 2006: 87.3 household, 81.4 adult[a]	Not addressed
Gutiérrez-Fisac, Regidor, and Alfaro (2010)	The 1986 Survey on Deficiencies, Disabilities and Handicaps The 1999 Survey on Disabilities and Health Conditions, the 2008 Survey on Disabilities, Personal self-sufficiency and Studies of dependence	Spain	Birth, 65+	1986–2007	Not addressed	Not addressed
Hashimoto et al. (2010)	The Comprehensive survey of Living Conditions 1995, 1998, 2001 and 2004	Japan	Birth, 65+	1995–2004	Not addressed	Not addressed
Moe and Hagen (2011)	The Norwegian Survey of Living Conditions	Norway	67-79, 80+	1987–2008	Vary between75%-65% and 68%-53%	Weights by age, sex.
Robine and Cambois (2013)	The European Statistics of Income and Living Conditions	EU(25)	65+	2005–2010	Vary. In 2010 between 57.3 and 94%.[b]	Not addressed
Cambois, Blachier, and Robine (2013)	Five French Population Health Surveys	France	50, 65	2003–2008	Vary between surveys (range 78%-91% at individual level)	No
Cutler, Ghosh, and Landrum (2014)	Medicare Current Beneficiary Survey + mortality data	USA	65+	1991–2008	Declining from 87.2 to 78% at initial interview, vary between 90-95% at follow up[a]	No
Jagger (2015)	UK decennial censuses	UK	Birth, 65+, 85+	2000/2002 2009/2011	Not addressed	Not addressed
Jeune et al. (2015)	SHARE	Denmark	65+	2004–2011	Vary between 61.5% and 69%	Weights
Solé-Auró and Alcañiz (2015)	The Catalan Health Survey	Catalonia	65+	1994–2011	Not addressed	Weights
Zimmer, Hidajat, and Saito (2015)	The Chinese Longitudinal Healthy Longevity Study	China	65+	2002–2011	96% for age 85+, 94-95% for age 65-79[a]	No
Crimmins, Zhang, and Saito (2016)	The National Health Interview Survey	USA	Birth, 65+	1970–2010	1998: 90% household, 83.8% adult, 2006: 87.3 household, 81.4 adult[a]	Weights by age, sex, race, ethnicity, metro-/non metro.
Freedman, Wolf, and Spillman (2016)	The National Long Term Care Survey (1982, 2004) and the National Health and Aging Trends Study (2011)	USA	65+	1982–2011	The NLTCS (above 95% in all waves)[d], NHATS: 71% in 2011[c]	Weights
Lagergren et al. (2016)	The Swedish Survey of Living Conditions	Sweden	65+	Mid 1980s -2006	Decreasing from 75 to 70%	No, but addressed + senstivity analysis
Sundberg et al. (2016)	The Swedish Panel Study of Living Conditions of the oldest Old, the Survey of Health, Ageing and Retirement in Europe	Sweden	77+	1992–2011	Vary between 84.4 and 95.4	Weights for part of sample
Steensma, Loukine, and Choi (2017)	The National Population Health Survey (1994–1999) The Canadian Community Health Survey (2000–2010)	Canada	20+, 65+	1994–2010	NPHS: 88.7%, 82.6% 89.7%, CCHS: declining from 84.7% to 72.3%	Not addressed
Kingston et al. (2017)	The Cognitive Ageing and Functioning Study I and II	England	65+	1991–2011	Declining from 80% to 56%	Weights
Strøreng et al. (2018)	The Nord-Trøndelag Health Surveys I,II, and III	Norway	30+	1984–1986, 2006–2008	Declining from 90% in HUNT I to 54% in HUNT III	No
Sperlich, Tetzlaff, and Geyer (2019)	The German Socio-Economic Panel	Germany	31-90	1995–2014	Not addressed	Weights

Notes: [a] Czajka and Beyler (2016), [b] The ESS Handbook for Quality Reports 2014. [c] NHATS Round 1 Sample design and selection. NHATS Technical Paper #1. [d] National Institute of Ageing, [e] Gu (2007).

References to the studies listed in Table A.1-1

Cai, L. and Lubitz, J. (2007). Was there a compression of disability for older Americans from 1992 to 2003? *Demography* 44(3): 479–495. doi:h10.1353/dem.2007.0022.

Cambois, E., Blachier, A., and Robine, J.-M. (2013). Aging and health in France: An unexpected expansion of disability among mid-adulthood over recent years. *European Journal of Public Health* 23(4): 575–581. doi:10.1093/eurpub/cks136.

Crimmins, E.M., Hayward, M.D., Hagedorn, A., Saito, Y., and Brouard, N. (2009). Change in disability-free life expectancy for Americans 70 years old and older. *Demography* 46(3), 627–646. doi:10.1353/dem.0.0070.

Crimmins, E.M. and Beltrán-Sánchez, H. (2010). Mortality and morbidity trends: Is there compression of morbidity? *The Journals of Gerontology*: Series B 66B(1): 75–86. doi:10.1093/geronb/gbq088.

Crimmins, E.M., Zhang, Y., and Saito, Y. (2016). Trends over 4 decades in disability-free life expectancy in the United States. *American Journal of Public Health* 106(7): 1287–1293. doi:10.2105/ajph.2016.303120.

Cutler, D.M., Ghosh, K., and Landrum, M.B. (2014). Evidence for significant compression of morbidity in the elderly U.S. population. In: Wise, D.A. (ed.). *Discoveries in the economics of ageing*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226146126.003.0002.

Czajka, J.L. and Beyler, A. (2016). Declining response rates in federal surveys: Trends and implications. Washington DC: Mathematica Policy Research.

Freedman, V.A., Wolf, D.A., and Spillman, C. (2016). Disability-free life expectancy over 30 Years: A growing female disadvantage in the US population. *American Journal of Public Health* 106(6): 1079–1085. doi:10.2105/AJPH.2016.303089.

Gu, D. (2007). General data quality assessment for the 2005 CLHLS wave. (CLHLS Technical Report No. 2007–01.) Durham, NC: Duke University Center for the Study of Aging and Human Development.

Gutiérrez-Fisac, J.L., Regidor, E., and Alfaro, M. (2010). Healthy life expectancies in Spain 1986–2007. Disability-free life expectancy and life expectancy in good perceived health in Spain and its autonomous communities. Madrid: Ministry of Health, Social Policy and Equality.

Hashimoto, S., Kawado, M., and Seko, R., Murakami, Y., Hayashi, M., Kato, M., Noda. T., Ojima, T., Nagai, M., and Tsuji, I. (2010). Trends in disability-free life expectancy in Japan, 1995–2004. *Journal of Epidemiology* 20(4): 308–312. doi:10.2188/jea.JE20090190.

Jagger, C. (2015). Trends in life expectancy and healthy life expectancy. Future of an ageing population: evidence review. London: Foresight, Government Office for Science.

Jeune, B. and Brønnum-Hansen, H. (2008). Trends in health expectancy at age 65 for various health indicators, 1987–2005, Denmark. *European Journal of Ageing* 5(4): 279–285. doi:10.1007/s10433-008-0100-x.

Jeune, B., Eriksen, M.L., Andersen-Ranberg, K., and Brønnum-Hansen, H. (2015). Improvement in health expectancy at ages 50 and 56 in Denmark during the period 2004–2011. *Scandinavian Journal of Public Health* 43(3): 254–259. doi:10.1177/1403494815569104.

Kingston, A., Wohland, P., Wittenberg, R., Robinson, L., Brayne, C., Matthews, F.E., and Jagger, C. (2017). A comparison of the Cognitive Function and Aging Studies (CFAS). *The Lancet* 390(10103): 1676–1684. doi:10.1016/S0140-6736(17)31575-1.

Lagergren, M., Johnell, K., Schön, P., and Danielson, M. (2016). Healthy life years in old age: Swedish development 1980–2010 according to different health indicators. *Scandinavian Journal of Public Health* 44(1): 55–61. doi:10.1177/1403494815605195.

Moe, J. and Hagen, T. (2011). Trends and variation in mild disability and functional limitations among older adults in Norway, 1986–2008. *European Journal of Aging* 8(1): 49–61. doi:10.1007/s10433-011-0179-3.

Robine, J-M. and Cambois, E. (2013). Healthy life expectancy in Europe. *Population and Societies* 499(4): 1–4. doi:10.3917/popsoc.499.0001.

Solé-Auró, A. and Alcañiz, M. (2015). Are we living longer but less healthy? Trends in mortality and morbidity in Catalonia (Spain), 1994–2011. *European Journal of Ageing* 12: 61–70. doi:10.1007/s10433-014-0317-9.

Sperlich, S., Tetzlaff, J., and Geyer, S. (2019). Trends in good self-rated health in Germany between 1995 and 2014: Do age and gender matter? *International Journal of Public Health* 64(6): 921–933. doi:10.1007/s00038-019-01235-y.

Steensma, C., Loukine, L., and Choi, B.C.K. (2017). Evaluating compression or expansion of morbidity in Canada: Trends in life expectancy and health-adjusted life expectancy

from 1994 to 2010. *Health Promotion and Chronic Disease Prevention in Canada* 37(3): 68–76. doi:10.24095/hpcdp.37.3.02.

Storeng, S., Krokstad, S., Westin, S., and Sund, E. (2018). Decennial trends and inequalities in healthy life expectancy: The HUNT Study, Norway. *Scandinavian Journal of Public Health* 46(1): 124–131. doi:10.1177/1403494817695911.

Sundberg, L., Agahi, N., Fritzell, J., and Fors, S. (2016). Trends in health expectancies among the oldest old in Sweden, 1992–2011. *The European Journal of Public Health* 26(6): 1069–1074. doi:10.1093/eurpub/ckw066.

Zimmer, Z., Hidajat, M., and Saito, Y. (2015). Changes in total and disability-free life expectancy among older adults in China: Do they portend a compression in morbidity? *International Journal of Population Studies* 1(1): 4–18. doi:10.18063/IJPS.2015.01.001.

A.2 The Nord-Trøndelag county

Nord-Trøndelag is a former county in the middle of Norway (since 2018 merged with Sør-Trøndelag county as Trøndelag county), consisting of 23 municipalities. See Figure A.2-1. Nord-Trøndealg is a rural and sparsely populated county, with an average income, educational attainment, and smoking prevalence a little lower than in the entire country (Holmen et al. 2003: 20). Its population aged 18+ years made up 2.97% in 1986, 2.87% in 1996, and 2.72% in 2007 of the total Norwegian population (Statistics Norway, Statbank Table 0745). The county does not include a city (Trondheim, the third largest city in Norway, is located in the neighbouring county Sør-Trøndelag). This is also a reason why it has a smaller share of immigrants and Norwegian borns with immigrant parents than the country as a whole (0.9% in 1987 and 3.2% in 2007 in Nord-Trøndelag versus 3.1% in 1987 and 8.8% in 2007 in Norway – Statistics Norway, Statbank Tables 05182 and 03037). However, in terms of geography, economy, industry, sources of income, age distribution, morbidity, and mortality, the county is regarded as fairly representative of Norway (Holmen et al., 2003: 20).

Figure A.2-1: The former Nord-Trøndelag county in Mid-Norway (left) and its municipality composition (right)

Source: Stina Aasen Lødemel/Allkunne CC BY-SA, kartgrunnlag Kartverket CC BY-SA 4.0 NO.

Table A.2-1 shows for each gender the expected remaining lifetime at age 65 years, in Norway and in Nord-Trøndelag. The figures are very close, except for a few months' discrepancy in the early 1980s.

Table A.2-1: Expected lifetime at age 65. Norway vs. Nord-Trøndelag county

		1981–1985	1986–1990	1991–1995	1996–2000	2001–2005	2006–2010
Nord-Trøndelag	men	15.1	14.9	15.1	16.0	16.8	17.7
	women	18.7	18.8	19.0	19.2	20.2	20.7
Norway	men	14.4	14.6	15.0	15.7	16.6	17.6
	women	18.5	18.7	19.0	19.5	20.1	20.8

Source: Statistics Norway, Statbank Table 05797.

Statistics Norway does not publish official life tables at the county level. However, it does publish population figures per gender and age groups at the county level (Statbank Table 07459) from 1986 onwards. Through private correspondence with Statistics Norway's Section for Population Statistics, I have received the number of deaths in Nord-Trøndelag per 5-year age group and gender for 1986 and 2007.[29] Based on this information, I have assembled the abridged life table for Nord-Trøndelag for these two years. The results for expected lifetime for age groups 50 to 54 years and higher are given in Table A.2-2, which also gives the corresponding figures for Norway in those years.

Table A.2-2: Expected lifetime in 1986 and 2007: Norway vs. Nord-Trøndelag county

	1986				2007			
	Women		Men		Women		Men	
age group	NT[a]	N[b]	NT	N	NT	N	NT	N
50–54	32.58	31.76	26.82	26.21	33.94	33.98	30.09	30.21
55–59	28.25	27.25	22.39	21.97	29.38	29.38	25.55	25.72
60–64	23.55	22.93	18.53	18.03	24.96	24.93	21.15	21.38
65–69	19.12	18.74	14.86	14.51	20.62	20.62	17.07	17.36
70–74	15.07	14.81	11.61	11.35	16.36	16.50	13.06	13.66
75–79	11.26	11.23	8.80	8.71	12.45	12.66	10.11	10.26
80–84	8.17	8.20	6.58	6.56	9.25	9.24	7.40	7.38
85+	5.86	5.82	4.58	4.80	6.83	6.83	4.89	4.87

Sources: [a]Nord-Trøndelag county, own computations based on Statistics Norway's Statbank Table 07459 and death figures provided by Statistics Norway's Section for population statistics [b]Norway: Human Mortality database.

The differences between the figures for Norway and Nord-Trøndelag county are minor except for the lower age groups in 1986. Because a certified life table for Norway is available for 1985, the year when the main part of the HUNT 1 Survey was conducted and 2007, because assembling the life table for Nord-Trøndelag table without information on

[29] I am grateful to Statistics Norway senior adviser Anders Stønsbø for these figures.

migration in and out of this county is likely to give a biased picture, and because there is a larger variance around death rates at the county level due to the small number of deaths, I use the life table for Norway for the analysis.

A.3 Construction of the FI indicator

Residents participating in the HUNT Surveys filled out the baseline questionnaire where one of the questions was "Do you suffer from any long-term illness or injury of a physical or psychological nature that impairs your functioning in your everyday life?" If answered "yes," the respondent was asked "Would you describe your impairment as slight, moderate or severe?" and could tick off the appropriate degree for the following specific impairments: motor impairment, vision impairment, hearing impairment, and impairment due to physical illness, impairment due to mental health problems. The baseline questionnaire had a paper format which made it possible for respondents to answer the specific questions without answering the main question. Of the HUNT1-respondents who answered affirmatively to the main impairment question, 98.2% also answered the specific impairment question (96.2% in HUNT2, 98.2% in HUNT3). At the same time, of the 404 respondents in HUNT1 who did not answer the main impairment question, 32.7% answered the specific impairment question (HUNT2: 51.4% of 5,575; HUNT3: 52.6% of 2,335).

A participant is classified as being functionally impaired if (1) the main impairment question is answered "yes," and (2) if not answered (missing) or answered "no," but any of the specific impairment questions is answered with "moderate" or "severe." On the other hand, a person is coded as disability-free ($FI = 0$) if (1) the main question is answered "no" or not answered (missing), and specific questions are answered only with "slight," and (2) the main question is answered "no" and answers to all specific questions are missing. The participant's disability status is coded as missing if the answer to both the main question and to all specific questions are missing. The rationale for this coding is to avoid participants being coded as disabled if they suffer from mild vision or hearing problems. In this respect, I am more conservative than Storeng et al. (2018), who also consider slight specific impairments as sufficient to warrant disability status. The coding rule for the outcome variable FI is given in Table A.3-1.

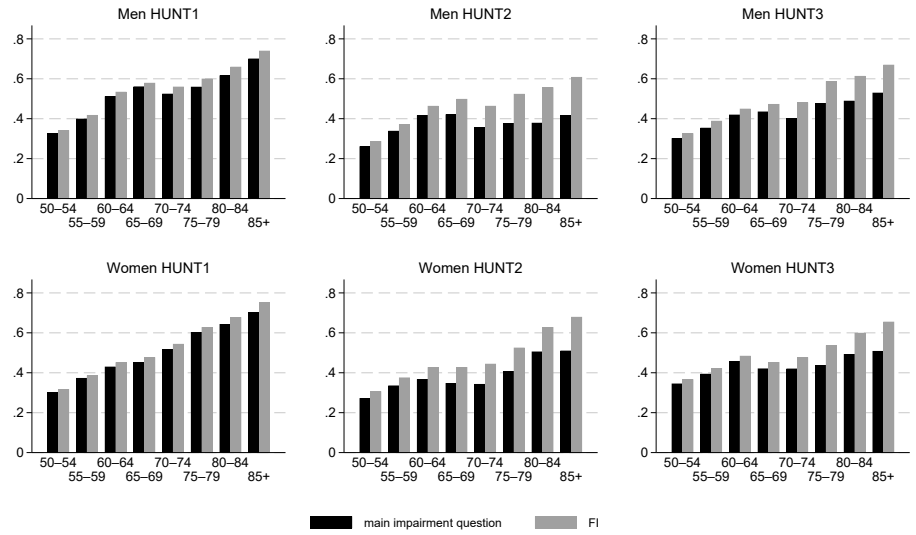
Figure A.3-1 shows the age profiles for FI (grey) as well as the corresponding profiles for affirmative answer to the main impairment question (black). The effect of taking "moderate" and "severe" answers to specific impairment questions into account is twofold: (1) It raises overall incidence of disability by a few percentage points, and (2) it eliminates or attenuates the non-monotonicity around ages 60 to 70 years present in the answer distribution to the main impairment question.

Table A.3-1: Coding of the outcome variable FI based on the answers to main and specific impairment questions

		Specific impairment questions		
		Not all missing, and none 'moderate' or 'severe'	At least one 'moderate' or 'severe'	All missing
main impairment question	'no'	0	1	0
	'yes'	1	1	1
	missing	0	1	missing

Note: 0 denotes disability-free; 1 denotes disabled.

Figure A.3-1: Age profiles for positive answer to the main impairment question (black) and for the FI variable (grey)



A.4 Descriptive statistics for the HUNT surveys

Table A.4-1 provides the descriptive statistics for the working sample.

Table A.4-1: Descriptive statistics. Nord-Trøndelag residents/participants aged 50+ years

Wave	Variable	Obs	Mean	St dev	Wave	Variable	Obs	Mean	St dev
1	male	39816	0.48	0.50	3	male	41255	0.477	0.500
	age	39816	66.31	10.26		age	41255	66.08	11.18
	married	36958	0.692	0.462		married	26921	0.692	0.462
	unmarried	36958	0.101	0.301		unmarried	26921	0.068	0.252
	divorced/separated	36958	0.026	0.160		divorced/separated	26921	0.098	0.297
	widowed	36958	0.181	0.385		widowed	26921	0.142	0.349
	SAH-Good[a]	36499	0.570	0.495		SAH-Good	25933	0.662	0.473
	FI[b]	36528	0.502	0.500		FI	26215	0.456	0.498
	BMI	35094	26.30	4.05		BMI	26701	27.60	4.23
	underweight	35094	0.010	0.100		underweight	26701	0.005	0.069
	overweight	35094	0.443	0.497		overweight	26701	0.477	0.499
	obese	35094	0.160	0.366		obese	26701	0.252	0.434
	diabetes	36438	0.056	0.229		diabetes	26918	0.065	0.246
	myocardial infarction	36241	0.053	0.225		myocardial infarction	26921	0.054	0.227
	angina pectoris	36264	0.093	0.290		angina pectoris	26915	0.060	0.237
	stroke[c]	36305	0.037	0.189		stroke	26918	0.040	0.195
	higher education	39816	0.045	0.206		higher education	41255	0.144	0.351
	education missing	39816	0.075	0.263		education missing	41255	0.029	0.167
	pe income[d]	36942	72.9	111.1		pe income	40115	149.0	191.8
	pe income > 0	19583	137.6	120.1		pe income > 0	23630	253.0	190.2
						total income	40188	278.1	175.9
2	male	39500	0.468	0.499					
	age	39500	66.68	11.37					
	married	38345	0.656	0.475					
	unmarried	38345	0.078	0.268					
	divorced/separated	38345	0.060	0.238					
	widowed	38345	0.205	0.404					
	SAH-Good	30073	0.593	0.491					
	FI	28437	0.430	0.495					
	BMI	29863	27.13	4.13					
	underweight	29863	0.006	0.080					
	overweight	29863	0.478	0.500					
	obese	29863	0.213	0.409					
	diabetes	30228	0.054	0.227					
	myocardial infarction	30200	0.066	0.248					
	angina pectoris	30165	0.102	0.303					
	stroke	30161	0.037	0.190					
	higher education	39500	0.088	0.284					
	education missing	39500	0.033	0.178					
	pe income	38287	88.6	131.7					
	pe income > 0	18760	180.9	136.7					
	total income[e]	38207	178.6	134.8					

Sources: HUNT Studies and Statistics Norway.
Notes: The variables SAH (self-assessed health), FI, diabetes, myocardial infraction, angina pectoris, and stroke are self-reported variables from the HUNT baseline questionnaire. Variable BMI is as measured at the health check. Remaining variables are from administrative records owned and merged to the HUNT data by Statistics Norway. [a] SAH-Good merges the answers "good" and "very good" to the SAH question. [b] *FI* is the FI indicator constructed as explained in Appendix A.3. [c] stroke records the answer to the question "Have you had or do you have stroke/brain haemorrhage?" [d,e] p(ension) e(ntitling) income and total income in 1000 NOK, converted to 2007 prices. 10.83 NOK maintained in 2007 the purchasing power of one euro in EU28. [e] total income is not available for HUNT1.

B. Technical appendix

The challenge is to estimate the prevalence rate of FI for a set of invitees to the survey – the population $\mathcal{P}$ – based on information about FI available for the set of respondents – the sample $\mathcal{S}$. I denote by $\#\mathcal{P}$ ($\#\mathcal{S}$) the number of persons in the population (the sample).

In this technical appendix, I explain the different steps to estimate $E(FI|A)$, the prevalence of FI in age group A, based on the information on FI for survey participants, (i.e., invitees for which the participation indicator R is 1). To simplify notation, I will from now on drop the conditioning on age group A throughout.

B.1 Estimator for the FI prevalence rate

If we observe only the value of the FI indicator for respondents, we can estimate $E(FI|R=1)$ but not $E(FI)$. In general, the relationship between $E(FI)$ and $E(FI|R=1)$ follows from the law of total probability:

$$E(FI) = E(FI|R=0)\Pr(R=0) + E(FI|R=1)\Pr(R=1).$$

The survey informs about $\Pr(R=1)$ (and thus about $\Pr(R=0)$) and about $E(FI|R=1)$, but not about $E(FI|R=0)$.

Assume now that the survey also informs about a vector of covariates w not only for respondents but for all invitees. Using the law of total expectation we have

$$E(FI) = \int_W E(FI|w)g(w)\mathrm{d}w, \tag{B.1}$$

where W is the domain of the covariates vector w, and $g(\cdot)$ is the density over this domain. The probability of the joint event that an invitee participates and shares the characteristics w is given by $g(R=1,w) = \Pr(R=1|w)g(w) = \Pr(R=1)g(w|R=1)$, where the second and third equalities follow from the definition of a conditional probability/density. Therefore, the right-hand side of (B.1) can be rewritten as

$$E(FI|A) = \int_W \underline{E(FI|w)} \times \left(\frac{\Pr(R=1|w)}{\Pr(R=1)}\right)^{-1} \times g(w|R=1)\mathrm{d}x. \tag{B.2}$$

In general, this right-hand side expression is still not operational to estimate $E(FI)$ because it involves the unobservable $E(FI|w)$.

Suppose first that observations are missing at random (MAR). This conveys the idea that once we have information about the background characteristics w, information about the outcome variable of interest, FI, does not matter to predict participation. Formally this means that $\Pr(R = 1|w, FI = 1) = \Pr(R = 1|w, FI = 0) = \Pr(R = 1|w)$. By 'inverting' these equalities busing Bayes' rule, we obtain $E(FI = 1|w, R = 1) = E(FI = 1|w, R = 0) = E(FI = 1|w)$.[30] Thus, MAR implies that once we condition on w, we may just as well rely on the prevalence rate of the sample, $E(FI = 1|w, R = 1)$, to make an inference about the prevalence rate of FI in the population, $E(FI = 1|w)$. In other words, the underlined term in (B.2) can be replaced by $E(FI|w, R = 1)$. The sample analogue of this term is the average of FI in the sample of respondents with characteristics w: $\frac{\sum_{i\in\mathcal{S}(w)} FI_i}{\#\mathcal{S}(w)}$. The sample analogue of $\Pr(R = 1|w)$ is the share of respondents with characteristics w in the set of invitees with those characteristics, $\frac{\#\mathcal{S}(w)}{\#\mathcal{P}(w)}$. Likewise, the overall response rate, $\frac{\#\mathcal{S}}{\#\mathcal{P}}$, is the sample analogue for $\Pr(R = 1)$. Because $g(w|R = 1)$ is the density of w among respondents, its sample analogue is the share of all respondents with those characteristics in the entire set of respondents, $\frac{\#\mathcal{S}(w)}{\#\mathcal{S}}$. Therefore, under MAR an estimator for (B.2) is

$$\sum_{w\in W} \frac{\sum_{i\in\mathcal{S}(w)} FI_i}{\#\mathcal{S}(w)} \times \left(\frac{\frac{\#\mathcal{S}(w)}{\#\mathcal{P}(w)}}{\frac{\#\mathcal{S}}{\#\mathcal{P}}} \right)^{-1} \times \frac{\#\mathcal{S}(w)}{\#\mathcal{S}},$$

or after simplifying,

$$\frac{1}{\#\mathcal{P}} \sum_{w\in X} \sum_{i\in\mathcal{S}(w)} FI_i \frac{\#\mathcal{P}(w)}{\#\mathcal{S}(w)}.$$

Denoting the probability that person i with characteristics w_i participates in the survey as $p(w_i)$, the previous expression can be rewritten as

$$\frac{1}{\#\mathcal{P}} \sum_{i\in\mathcal{S}} FI_i \frac{1}{p(w_i)}.$$

Finally, the number of individuals in the population can be written as

[30] This property is also called 'conditional independence'. Recall that for a binary variable, its expectation coincides with the probability of that variable taking the value 1.

$$\#\mathcal{P} = \sum_{w \in W} \#\mathcal{P}(w) = \sum_{w \in W} \frac{\#\mathcal{S}(w)}{\frac{\#\mathcal{S}(w)}{\#\mathcal{P}(w)}} = \sum_{w \in W} \sum_{i \in \mathcal{S}(w)} \frac{1}{\frac{\#\mathcal{S}(w)}{\#\mathcal{P}(w)}} = \sum_{i \in \mathcal{S}} \frac{1}{p(w_i)},$$

where $p(w)$ is the probability that a person with characteristics w participates in the survey – this can be obtained by estimating a probit model for participation (cf below).

Substituting the right-hand side for $\#P$ in the previous expression finally gives

$$\widehat{E(FI)}_{MAR} = \sum_{i \in \mathcal{S}} FI_i \frac{\frac{1}{p(w_i)}}{\sum_{j \in \mathcal{S}} \frac{1}{p(w_j)}},$$

which is the inverse-probability weighted average of FI for survey participants.

As argued in the main text, MAR is a strong assumption, especially in the context of health surveys. For example, FI itself may reduce the likelihood that an invitee participates in the survey, in which case we have that $\Pr(FI = 1|w, R = 1) < \Pr(FI = 1|w) < \Pr(FI = 1|w, R = 0)$. Thus, replacing the underlined term in (B.2) by $E(FI = 1|w, R = 1)$ is going to result in an underestimate of $E(FI = 1)$.

To obtain a consistent estimate of $E(FI = 1|w)$ without the MAR assumption, we need to formulate a joint model for FI and the participation decision. The standard model in the selection literature is a variant of the one proposed by Heckman (1976):

$$\begin{aligned} R_i &= I(R_i^* = w_i'a + u_i > 0), \\ FI_i &= I(FI_i^* = w_i'b + e_i > 0), \\ FI_i &\text{ observed if and only if } R_i = 1, \\ \begin{pmatrix} u_i \\ e_i \end{pmatrix} &\sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}), \end{aligned} \tag{B.3}$$

where $I(\cdot)$ is the indicator function (i.e., $I(s) = 1(0)$ if s is true (false)), a and b are parameter (column) vectors with the same dimension as w,[31] and r is the correlation coefficient between the error terms u and e. Moreover, a and b are so defined such that $cov(w, u) = cov(w, e) = 0$.[32] Let $(\widehat{a}', \widehat{b}', \widehat{r})$ be the (row) vector of maximum likelihood

[31] I assume w includes a constant term.

[32] Thus, I do not attach any interpretation of causality to any component of b: b_k may capture the direct effect of variable w_{ki} on FI_i^* but may in addition include the expected effect of unobserved variables with which w is correlated. Formally, let $FI^* = w'\widetilde{b} + v'c$ where v is a vector of unobserved variables and $\widetilde{b}$ is the causal effect of the observables w on FI^*. Then define e as $v'c - E(v|w)'c$. Then $FI^* = w'\widetilde{b} +$

estimates of the model. Then, a consistent estimate for the probability of participation for invitee i with characteristics vector w_i is $p(w_i'\widehat{a}) \stackrel{\text{def}}{=} \Phi(w_i'\widehat{a})$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Likewise, a consistent estimate for the probability that this individual is functionally impaired is $\widehat{E(FI|w_i)} \stackrel{\text{def}}{=} \Phi(w_i'\widehat{b})$.

Thus, if the MAR assumption does not hold, we can obtain a consistent estimate of $E(FI|w)$ by estimating model (B.3), and in a next step obtain an estimate for $E(FI)$ either by (1) imputing $\Phi(w_i'\widehat{b})$ to every member of the population and then estimating the prevalence rate for the population as

$$\frac{\sum_{i\in\mathcal{P}} \Phi(w_i'\widehat{b})}{\#\mathcal{P}},$$

('imputation'); or (2) imputing $\Phi(x_i'\widehat{b})$ to every member of the sample (i.e., to every *participating* member of the population) and then estimating the prevalence rate for the population as

$$\sum_{i\in\mathcal{S}} \Phi(w_i'\widehat{b}) \frac{\frac{1}{p(w_i'\widehat{a})}}{\sum_{j\in\mathcal{S}} \frac{1}{p(w_j'\widehat{a})}}$$

('inverse-probability weighting').

B.2 Estimation issues of the Heckman model without an exclusion restriction

Both equations in (B.3) contain the same vector w. In principle, the parameter vector b is identified due to the distributional assumptions of the model. However, in practice, the likelihood function is very flat in the neighbourhood of the true parameter vector b, and the intuition for this was given in the main text. The standard solution to secure identification is the inclusion of a variable which is strongly correlated with R^* but uncorrelated with FI^* (see Cameron and Trivedi 2005: 551). Such an exclusion instrument is rarely present in health surveys.

However, identification of b can also be obtained by imposing a restriction on r. Altonji, Elder, and Taber (2002, 2005) show that if the mechanism determining which variables in a survey end up on the list of observable characteristics and which variables end up as unobservables is random, then the degree of selection on unobservables (SoU) will

$E(v|w)'c + e$. Suppose that $E(v|w) = \Gamma w$, where Γ is a matrix with appropriate dimensions. Then substitution gives $FI^* = w'\widetilde{b} + w'\Gamma'c + e$. Defining b as $\widetilde{b} + \Gamma'c$ then results in the second line of (B.3). Because $E(e|w) = E[v'c - E(v|w)'c|w] = 0$, it follows that $E(we) = cov(w, e) = 0$. For the same reason, I define a such that $cov(w, u) = 0$.

be as strong as the degree of selection on observables (SoO). To formalise this 'SoU=SoO' restriction, consider the linear projection of the latent participation variable R_i^* on $w_i'b$ ('the observables') and e_i ('the unobservables'):

$$\text{Proj}(R^*|w'b, e) = \phi_0 + \phi_{w'b} w'b + \phi_e e.$$

The 'SoU=SoO' restriction then amounts to $\phi_e = \phi_{w'b}$. Because $\phi_y = \frac{cov(R^*,y)}{var(y)}$ $(y = w'b, e)$,[33] the distributional assumptions on (u, e), together with the orthogonality conditions, imply that $\phi_e = \rho$ and $\phi_{w'b} = \frac{cov(w'a,w'b)}{var(w'b)}$. Therefore the 'SoU=SoO' restriction amounts to

$$\rho = \frac{cov(w'a, w'b)}{var(w'b)}. \tag{B.4}$$

This is a non-linear restriction on the parameters of interest, ρ and $\binom{a}{b}$. Maximum likelihood estimation of (B.3) subject to this restriction can easily be implemented through an iterative procedure, as explained in Section B.3.

So far, I have assumed that the variables that make it onto the list of observables are a random selection of all variables that explain FI^*. What if some of the variables destined for the list of observables nevertheless do not make it to that list – for example, because only their lagged values can be measured?

To address this question, I partition the $(n \times 1)$ vector w into a vector of variables containing information about health and socio-demographic characteristics from the previous wave (x, dimensions $(n_x \times 1)$) and variables with information about more recent health shocks that occurred after the previous wave (z, dimensions $(n_z \times 1)$). So $w' = (x', z')$, and likewise $a' = (a_x', a_z')$ and $b' = (b_x', b_z')$. By definition, z is not observable for non-participants. What are the consequences of such non-random exclusion from the initial variable vector w for the degree of selection on unobservables versus the degree of selection on observables? Because z is unobservable, model (B.3) can no longer be estimated. Let

$$z = \Delta x + \eta, \text{ with } E\eta = 0 \text{ and } E\eta\eta' = V_\eta, \tag{B.5}$$

[33] See Sargent (1976: ch II) for an introduction to linear projections. In general, the linear projection of variable y on the vector x is given by $[E(xx')]^{-1}E(xy)$. If the variables making up vector x are uncorrelated with each other, $E(xx')$ becomes a diagonal matrix, and the ith element of the linear projection simplifies to $\frac{cov(y,x_i)}{var(x_i)}$. In the present setup, $w'b$ will indeed be uncorrelated with e by the way the model is defined (see the previous footnote).

where Δ is an $(n_z \times n_x)$-coefficient matrix, and η is a $(n_z \times 1)$ vector with 'innovations' – new unpredictable health shocks – with covariance matrix V_η. Further assume that $Ex\eta' = 0$.[34]

In model (B.3), the part of the unobservable z that is correlated with x will be picked up by x with a new coefficient vector α (β) in the equation for R^* (FI^*), while the part of z that is uncorrelated with x will together with the original error terms u and e form new error terms ν and ε, which have a new covariance ρ. This results in the following reduced model that can be estimated:

$$\begin{gathered} R_i = I(R_i^* = x_i'\alpha + \nu_i > 0), \\ FI_i = I(FI_i^* = x_i'\beta + \varepsilon_i > 0), \\ FI_i \text{ observed if and only if } R_i = 1, \text{ and} \\ \begin{pmatrix}\nu_i \\ \varepsilon_i\end{pmatrix} \sim N(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}var(\nu) & \rho \\ \rho & var(\varepsilon)\end{pmatrix}), \end{gathered} \tag{B.6}$$

where $\alpha \stackrel{\text{def}}{=} a_x + \Delta' a_z$, $\beta \stackrel{\text{def}}{=} b_x + \Delta' b_z$, $\nu \stackrel{\text{def}}{=} u + \eta' a_z$, and $\varepsilon \stackrel{\text{def}}{=} e + \eta' b_z$, implying that $var(\nu) = 1 + a_z' V_\eta a_z$, $var(\varepsilon) = 1 + b_z' V_\eta b_z$, and $\rho \stackrel{\text{def}}{=} cov(\nu, \varepsilon) = r + a_z' V_\eta b_z$.

For this reduced model, the linear projection of R^* on x and ε is given by

$$\text{Proj}(R^* | x'\beta, \varepsilon) = \phi_{00} + \phi_{x'\beta} x'\beta + \phi_\varepsilon \varepsilon,$$

with

$$\phi_{x'\beta} = \frac{cov(x'\alpha, x'\beta)}{var(x'\beta)} \text{ and } \phi_\varepsilon = \rho.$$

We then have the following result.

Theorem Suppose that for the original model (B.3) the assumptions on the variable selection process of Altonji, Elder, and Taber (2005) are satisfied such that (B.4) holds with reference to the vector of characteristics w. Suppose that the subvector z is removed from the list of observables $w' = (x', z')$ such that the reduced model becomes (B.6). Finally, suppose that the relation between vector z and vector x is given by (B.5). Then ϕ_ε can be written as a weighted average of $\phi_{x'\beta}$ and $\frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)}$:

[34] This assumption can be justified by a similar reasoning as in footnote 32.

$$\phi_\varepsilon = \phi_{x'\beta} \times (1-\omega) + \frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)} \times \omega, \tag{B.7}$$

with $\omega \stackrel{\text{def}}{=} \frac{var(\eta' b_z) + \frac{var(\eta' b_z)}{var(w'b)}}{1+var(\eta' b_z)} \in [0,1]$. Therefore, $\phi_\varepsilon \gtrless \phi_{x'\beta}$ if and only if $\phi_{x'\beta} \lessgtr \frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)}$. Because $\phi_{x'\beta} = \frac{cov(x'\alpha, x'\beta)}{var(x'\beta)}$, this inequality can also be stated as $\frac{cov(w'a, w'b)}{var(w'b)} \lessgtr \frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)}$.

Proof The original projection of R^* onto $(1, w'b, e)$ is $\phi_0 + \phi_{w'b} w'b + \phi_e e$ with $\phi_{w'b} = \frac{cov(w'a, w'b)}{var(w'b)}$ and $\phi_e = \frac{cov(u,e)}{var(e)} = r$. In the reduced model, the linear projection of R^* onto $(1, x'\beta, \varepsilon)$ is $\phi_{00} + \phi_{x'\beta} x'\beta + \phi_\varepsilon \varepsilon$ with $\phi_{x'\beta} = \frac{cov(x'\alpha, x'\beta)}{var(x'\beta)}$ and $\phi_\varepsilon = \frac{cov(\nu,\varepsilon)}{var(\varepsilon)} = \frac{r + cov(\eta' a_z, \eta' b_z)}{1+var(\eta' b_z)}$.

Because $cov(w'a, w'b) = cov(x'\alpha, x'\beta) + cov(\eta' a_z, \eta' b_z)$ and likewise $var(w'b) = var(x'\beta) + var(\eta' b_z)$, we can write $\phi_{x'\beta}$ as

$$\begin{aligned} \phi_{x'\beta} &= \frac{cov(w'a, w'b) - cov(\eta' a_z, \eta' b_z)}{var(w'b) - var(\eta' b_z)}, \text{ or} \\ &= \frac{r\ var(w'b) - cov(\eta' a_z, \eta' b_z)}{var(w'b) - var(\eta' b_z)}, \end{aligned} \tag{B.8}$$

where the second equality follows from the assumption that $\phi_e = \phi_{w'b}$ or $r = \frac{cov(w'a, w'b)}{var(w'b)}$. Rearranging this then gives

$$r = \frac{\phi_{x'\beta} \times [var(w'b) - var(\eta' b_z)] + cov(\eta' a_z, \eta' b_z)}{var(w'b)}.$$

Because $\phi_\varepsilon = \frac{r + cov(\eta' a_z, \eta' b_z)}{1+var(\eta' b_z)}$, it follows that

$$\phi_\varepsilon = \frac{\frac{\phi_{x'\beta} \times [var(w'b) - var(\eta' b_z)] + cov(\eta' a_z, \eta' b_z)}{var(w'b)} + cov(\eta' a_z, \eta' b_z)}{1 + var(\eta' b_z)}.$$

Collecting the terms with $\phi_{x'\beta}$ and $cov(\eta' a_z, \eta' b_z)$ then gives (B.7). Using (B.8), the inequality $\phi_{x'\beta} \lesseqgtr \frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)}$ can be written as

$$\frac{cov(w'a, w'b) - cov(\eta' a_z, \eta' b_z)}{var(w'b) - var(\eta' b_z)} \lesseqgtr \frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)},$$

which can be simplified to

$$\frac{cov(w'a, w'b)}{var(w'b)} \lesseqgtr \frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)}. \quad \blacksquare$$

The theorem supports the intuitive idea that if in the original model selection on unobservables is as strong as on observables, then in the reduced model, selection will be stronger on unobservables than on observables if and only if $\frac{cov(x'\alpha, x'\beta)}{var(x'\beta)} < \frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)}$, that is, if selection on the recent health shocks η is stronger than on the 'older' observables x. The theorem has two interesting implications that further underscore this intuition.

Corollary 1 If $var(\eta) = 0_{n_z \times n_z}$ (i.e., if the vector z is perfectly predictable by the observable vector x), then $\phi_\varepsilon = \phi_{x'\beta}$, (i.e., in the reduced model, selection on unobservables will also be as strong as on observables).

Proof This follows from that fact that ω approaches 0 and $\frac{cov(\eta' a_z, \eta' b_z)}{var(\eta' b_z)}$ approaches a finite number when $var(\eta) \to 0_{n_z \times n_z}$. $\blacksquare$

Intuitively, if $var(\eta) = 0_{n_z \times n_z}$, then all information in z is already contained in x. Therefore, not observing z does not affect the strengths of selection.

Corollary 2 If $\Delta = 0_{n_z \times n_x}$, (i.e., if the vector z consists of only shocks that are unpredictable by the observable vector x), then $\phi_\varepsilon \gtreqless \phi_{x'\beta}$ if and only if $\frac{cov(x' a_x, x' b_x)}{var(x' b_x)} \lesseqgtr \frac{cov(z' a_z, z' b_z)}{var(z' b_z)}$.

Proof If $\Delta = 0$, then $z = \eta$. Working out the inequality $\frac{cov(w'a, w'b)}{var(w'b)} \gtreqless \frac{cov(z' a_z, z' b_z)}{var(z' b_z)}$ gives $\frac{cov(x' a_x, x' b_x)}{var(x' b_x)} \lesseqgtr \frac{cov(z' a_z, z' b_z)}{var(z' b_z)}$. $\blacksquare$

To fix these ideas, suppose that $\frac{cov(x' a_x, x' b_x)}{var(x' b_x)} < \frac{cov(z' a_z, z' b_z)}{var(z' b_z)}$, so that in the original model, selection on the 'older' x variables is weaker than on the more recent z variables. Next, the latter variables are removed from the vector w and lumped together with the error term e. Then selection on the x variables will be weaker than selection on the new composed error term $e + z' b_z$.

Finally, a comment on the normalisation of the reduced model (B.6) is required. In the original model (B.3), the variances of u and e were normalised to 1. This is because

R^* and FI^* are latent variables that are not observable to the researcher; only their sign is observed. This means that the variances of the two error terms cannot be identified and hence these are normalised to 1.

In the reduced model (B.6), these variances are no longer unified. If the reduced model is taken to the data, an analogous normalisation needs to take place. This is done by dividing the expressions for R^* and FI^* by the square root of the variance of u and e , respectively. Hence, we get

$$R_i = I(\frac{R_i^*}{\sqrt{1+a_z' V_\eta a_z}} = x_i' \frac{\alpha}{\sqrt{1+a_z' V_\eta a_z}} + \frac{\nu_i}{\sqrt{1+a_z' V_\eta a_z}} > 0),$$
$$FI_i = I(\frac{FI_i^*}{\sqrt{1+b_z' V_\eta b_z}} = x_i' \frac{\beta}{\sqrt{1+b_z' V_\eta b_z}} + \frac{\varepsilon_i}{\sqrt{1+b_z' V_\eta b_z}} > 0).$$

The result is the estimatable reduced model in terms of the new latent variables and new coefficients (both with superscript n):

$$\begin{gathered} R_i = I(R_i^{n*} = x_i'\alpha^n + \nu_i^n > 0), \\ FI_i = I(FI_i^{n*} = x_i'\beta^n + \varepsilon_i^n > 0), \\ FI_i \text{ observed if and only if } R_i = 1, \text{ and} \\ \begin{pmatrix} \nu_i^n \\ \varepsilon_i^n \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^n \\ \rho^n & 1 \end{pmatrix}, \end{gathered} \tag{B.9}$$

where $\rho^n = \frac{r + a_z' V_\eta b_z}{\sqrt{1+a_z' V_\eta a_z}\sqrt{1+b_z' V_\eta b_z}}$.

The projection of R^{n*} on $x'\beta^n$ and ε^n has coefficients

$$\begin{aligned} \phi_{x'\beta^n} &= \frac{cov(R^{n*}, x'\beta^n)}{var(x'\beta^n)} = \frac{cov(x'\alpha^n, x'\beta^n)}{var(x'\beta^n)} = \frac{\sqrt{1+b_z' V_\eta b_z}}{\sqrt{1+a_z' V_\eta a_z}} \frac{cov(x'\alpha, x'\beta)}{var(x'\beta)} \\ &= \frac{\sqrt{1+b_z' V_\eta b_z}}{\sqrt{1+a_z' V_\eta a_z}} \phi_{x'\beta}, \text{ and} \\ \phi_{\varepsilon^n} &= \frac{cov(R^{n*}, \varepsilon^n)}{var(\varepsilon^n)} = \frac{cov(u^n, \varepsilon^n)}{var(\varepsilon^n)} = \frac{r + a_z' V_\eta b_z}{\sqrt{1+a_z' V_\eta a_z}\sqrt{1+b_z' V_\eta b_z}} \\ &= \frac{\sqrt{1+b_z' V_\eta b_z}}{\sqrt{1+a_z' V_\eta a_z}} \phi_\varepsilon. \end{aligned}$$

It then follows that $\phi_{\varepsilon^n} \gtrless \phi_{x'\beta^n}$ if and only if $\phi_{\varepsilon} \gtrless \phi_{x'\beta}$. I have therefore shown that any statement on the relation between ϕ_{ε} and $\phi_{x'\beta}$ carries over to the relation between ϕ_{ε^n} and $\phi_{x'\beta^n}$. The econometric results presented in the main text are for the normalised reduced model (B.9), but with the superscript ^{n} dropped for convenience.

B.3 Description of the iterative algorithm to estimate Heckman probit model under the AET identification restriction

The H-AET model is the bivariate selection model (B.9) with the restriction that $\rho = \frac{cov(x'\alpha, x'\beta)}{var(x'\beta)}$. It is estimated using the `heckprobit` command in Stata 14 under the restriction that $\widehat{\rho}_s = \frac{cov(x'\widehat{\alpha}_{s-1}, x'\widehat{\beta}_{s-1})}{var(x'\widehat{\beta}_{s-1})}$, where $(\widehat{\alpha}_{s-1}, \widehat{\beta}_{s-1})$ is the estimated parameter vector from iteration $s-1$.[35] For the first iteration ($s = 1$) `heckprobit` is executed without any restriction. Convergence is declared when $|\widehat{\rho}_s - \widehat{\rho}_{s-1}| < 0.00001$. Standard errors are obtained by bootstrapping. For each estimation, 100 new samples are drawn. The bth sample, S^b, yields the estimated coefficient vector $(\widehat{\alpha}^b, \widehat{\beta}^b, \widehat{\rho}^b)$. The standard error of any estimated coefficient is then the standard deviation across these 100 point estimates. Tables 3 and 4 in the main text report the marginal effects. The corresponding standard errors are also based on bootstrapping the sample.

B.4 Details about the implementation procedure of Section 4.2

Recall that the vector of explanatory variables for HUNT3, x_3, consists of variables measured during HUNT1, x_1, and the indicator of functional impairment measured during HUNT2, FI_2, as well as register data on education and income in HUNT3. The vector of explanatory variables for HUNT2, x_2, consist of variables measured during HUNT1 and register data on education and income in HUNT2.

Because I condition on participation in HUNT1 and because the register data on education and income (whether during HUNT3 or HUNT2) are observable anyway due to participation in HUNT1, I subsume these data in the vector x_1. Let X_1 be the set of possible vectors x_1.

Therefore the characteristics space in HUNT2 is $X_2 = X_1$, whereas that for HUNT3 is $X_3 = X_1 \times \{0, 1\}$, where $\{0, 1\}$ is the set of possibilities for FI_2. In the main text, I have shown the estimation results for model (8) [or (B.9)] for HUNT2 and for HUNT3. Let the list of estimated coefficients for HUNT2 be given by $(\widehat{\alpha}_2', \widehat{\beta}_2', \widehat{\rho}_2)$, and that for the model for HUNT3 by $(\widehat{\alpha}_3', \widehat{\beta}_3', \widehat{\rho}_3)$, and to make clear that the model for HUNT3 also uses FI_2 as an explanatory variable, I partition $\widehat{\alpha}_3'$ as $(\widehat{\alpha}_3^{-\prime}, \widehat{\alpha}_3^{FI_2})$ and $\widehat{\beta}_3'$ as $(\widehat{\beta}_3^{-\prime}, \widehat{\beta}_3^{FI_2})$.

[35] $cov(\cdot)$ and $var(\cdot)$ refer here to the estimation sample covariance and variance.

I am interested in estimating $E(FI_3)$. I first condition on participation. Recall that the 'population' is considered to consist of all participants in HUNT1 who survive until HUNT3. Given participation in HUNT1, there are three possible participation sequences: $(R_2 = 1, R_3 = 1)$, $(R_2 = 1, R_3 = 0)$, and $(R_2 = 0, R_3 = 0)$. The fourth possibility, $(R_2 = 0, R_3 = 1)$, is ruled out by the assumption of monotone attrition.

Conditioning on participation in HUNT2 or not, we get from the law of total probability that

$$E(FI_3) = E(FI_3|R_2 = 1)\Pr(R_2 = 1) + E(FI_3|R_2 = 0)\Pr(R_2 = 0).$$

I now expand each of the two conditional expectations further by conditioning on x_3. Then,

$$E(FI_3|R_2 = 1) = \int_{X_3} E(FI_3|x_3, R_2 = 1)g\left(x_3|R_2 = 1\right)\mathrm{d}x_3,$$

and

$$E(FI_3|R_2 = 0) = \int_{X_3} E(FI_3|x_3, R_2 = 0)g\left(x_3|R_2 = 0\right)\mathrm{d}x_3.$$

The sample analogue of $E(FI_3|R_2 = 1)$ is given by[36]

$$\sum_{x_3 \in X_3} \frac{\sum_{i \in S_3(x_3)} \Phi(x'_{3i}\widehat{\beta}_3) + \sum_{i \in S_2(x_3) \setminus S_3(x_3)} \Phi(x'_{3i}\widehat{\beta}_3)}{\#S_2(x_3)} \times \frac{\#S_2(x_3)}{\#S_2}, \tag{B.10}$$

where $S_2(x_3) = S_2(x_1, FI_2)$ is the collection of individuals (in a particular age and gender group) with characteristics x_1 in HUNT1 and FI indicator FI_2 in HUNT2. (B.10) can be simplified to

$$\frac{\sum_{x_3 \in X_3} \sum_{i \in S_2(x_3)} \Phi(x'_{3i}\widehat{\beta}_3)}{\#S_2},$$

or

$$\widehat{E(FI_3|R_2 = 1)} = \frac{\sum_{i \in S_2} \Phi(x'_{3i}\widehat{\beta}_3)}{\#S_2}. \tag{B.11}$$

[36] Refer to the discussion after (B.2).

That is, we impute $\Phi(x'_{3i}\widehat{\beta}_3)$ to each person participating in HUNT2, irrespective of whether he or she participates in HUNT3 or not, and then take the average to get the prevalence rate of FI in HUNT3.

However, we can also rewrite (B.10) as

$$\sum_{x_3 \in X_3} \left(\underline{\frac{\sum_{i \in S_3(x_3)} \Phi(x'_{3i}\widehat{\beta}_3)}{\#S_3(x_3)}} \frac{\#S_3(x_3)}{\#S_2(x_3)} + \underline{\frac{\sum_{i \in S_2(x_3) \setminus S_3(x_3)} \Phi(x'_{3i}\widehat{\beta}_3)}{\#S_2(x_3) - \#S_3(x_3)}} \frac{\#S_2(x_3) - \#S_3(x_3)}{\#S_2(x_3)} \right) \times \frac{\#S_2(x_3)}{\#S_2}.$$

Because $\Phi(x'_3\widehat{\beta}_3)$ is just a probability that depends on x_3, the two underlined terms are almost identical (strictly speaking, they will be identical in their probability limit as the number of observations goes to infinity), and the expression simplifies to

$$\sum_{x_3 \in X_3} \frac{\sum_{i \in S_3(x_3)} \Phi(x'_{3i}\widehat{\beta}_3)}{\#S_3(x_3)} \times \frac{\#S_2(x_3)}{\#S_2}, \tag{B.12}$$

which can also be rewritten as

$$\sum_{x_3 \in X_3} \sum_{i \in S_3(x_3)} \Phi(x'_{3i}\widehat{\beta}_3) \frac{1}{\frac{\#S_3(x_3)}{\#S_2(x_3)}} \times \frac{1}{\#S_2}, \text{ or}$$

$$\sum_{i \in S_3} \Phi(x'_{3i}\widehat{\beta}_3) \frac{1}{p(x'_{3i}\widehat{\alpha}_3)} \times \frac{1}{\#S_2}, \tag{B.13}$$

where $p(x'_{3i}\widehat{\alpha}_3) = \Phi(x'_{3i}\widehat{\alpha}_3)$ is the probability of participating in HUNT3, given participation in HUNT2. As a last step, we can note that

$$\#S_2 = \sum_{x_3 \in X_3} \#S_2(x_3) = \sum_{x_3 \in X_3} \frac{\#S_3(x_3)}{\frac{\#S_3(x_3)}{\#S_2(x_3)}} = \sum_{x_3 \in X_3} \sum_{i \in S_3(x_3)} \frac{1}{\frac{\#S_3(x_3)}{\#S_2(x_3)}} = \sum_{j \in S_3} \frac{1}{p(x'_{3j}\widehat{\alpha}_3)},$$

so that

$$\widehat{E(FI_3|R_2 = 1)} = \sum_{i \in S_3} \Phi(x'_{3i}\widehat{\beta}_3) \frac{\frac{1}{p(x'_{3i}\widehat{\alpha}_3)}}{\sum_{j \in S_3} \frac{1}{p(x'_{3j}\widehat{\alpha}_3)}}, \tag{B.14}$$

which is the inverse-probability weighted average of the imputed degree of FI among HUNT3 participants, unconditional on participation in HUNT3, but conditional on having participated in HUNT2. Thus we have two equivalent ways to estimate $E(FI_3|R_2=1)$: Either impute $\Phi(x'_{3i}\widehat{\beta}_3)$ to all invitees to HUNT3 (but excluding the (re-)joiners – cf monotone attrition assumption), or impute $\Phi(x'_{3i}\widehat{\beta}_3)$ to the participants in HUNT3 and use inverse-probability weighting to make this HUNT3 sample representative for HUNT2.

This leaves the task of estimating $E(FI_3|\text{permanent attritor})$. Recall that $x_3 = (x_1, FI_2)$. The expectation can be expanded as

$$
\begin{aligned}
&\int_{X_1\times\{0,1\}} E(FI_3|x_1, FI_2, \text{pa}) g\left(x_1, FI_2|\text{pa}\right) \mathrm{d}\left(x_1, FI_2\right) \\
&= \int_{X_1} E(FI_3|x_1, FI_2=1, \text{pa}) g\left(x_1, FI_2=1|\text{pa}\right) \mathrm{d}x_1 + \\
&+ \int_{X_1} E(FI_3|x_1, FI_2=0, \text{pa}) g\left(x_1, FI_2=0|\text{pa}\right) \mathrm{d}x_1 \\
&= \int_{X_1} \{E(FI_3|x_1, FI_2=1, \text{pa}) \Pr(FI_2=1|x_1, \text{pa}) + \\
&+ E(FI_3|x_1, FI_2=0, \text{pa})[1-\Pr(FI_2=1|x_1, \text{pa})]\} g\left(x_1|\text{pa}\right) \mathrm{d}x_1,
\end{aligned}
$$

where pa stands for permanent attritor.

I propose (1) to estimate $E(FI_3|x_1, FI_2=j,\text{pa})$ by means of the predicted marginal probability of having functional impairment, using the estimated model for HUNT3, i.e., $\widehat{E(FI_3|x_1, FI_2}=j) = \Phi(\widehat{\beta}_{3FI}\times j, x_{3i}^{-\prime}\widehat{\beta}_3^-)$, $j=0,1$; and (2) to estimate $\Pr(FI_2=1|x_1,\text{pa})$ by the ratio of the estimated probability that an individual invited to HUNT2 does not participate and is functionally impaired ($\Pr(FI_2=1, R_2=0|x_1)$) to the estimated probability that the same individual does not participate in HUNT2 ($\Pr(R_2=0|x_1)$), and by assumption becomes a permanent attritor: $\widehat{\Pr(FI_2=1|x_1, pa)} = \widehat{\Pr(FI_2=1|x_1, R_2=0)} = \frac{\Phi_B(x'_{1i}\widehat{\beta}_2, -x'_{1i}\widehat{\alpha}_2, -\widehat{\rho}_2)}{1-\Phi\left(x'_{1i}\widehat{\alpha}_2\right)}$.[37]

[37] $\Pr(FI=1, R=0) = \Pr(x'\beta+\varepsilon>0, x'\alpha+\nu<0) = \Phi_B(x'\beta, -x'\alpha, -\rho)$, and $\Pr(R=0) = \Pr(x'\alpha+\nu<0) = 1-\Pr(x'\alpha+\nu>0) = \Phi(x'\alpha)$, where $\Phi_B(\cdot,\rho)$ is the cumulative distribution function of the bivariate standard normal distribution with correlation ρ, and $\Phi(\cdot)$ is the cumulative distribution function of the univariate standard normal distribution. ρ is the correlation coefficient between the error terms – see (8) or (B.9).

Thus, defining

$$\pi(x_{1i}) \stackrel{\text{def}}{=} \Phi(\widehat{\beta}_{3FI} \times 1, x_{3i}^{-\prime}\widehat{\beta}_{3}^{-}) \times \frac{\Phi_B(x_{1i}'\widehat{\beta}_2, -x_{1i}'\widehat{\alpha}_2, -\widehat{\rho}_2)}{1-\Phi\left(x_{1i}'\widehat{\alpha}_2\right)} + \Phi(\widehat{\beta}_{3FI} \times 0, x_{3i}^{-\prime}\widehat{\beta}_{3}^{-}) \times \left(1 - \frac{\Phi_B(x_{1i}'\widehat{\beta}_2, -x_{1i}'\widehat{\alpha}_2, -\widehat{\rho}_2)}{1-\Phi\left(x_{1i}'\widehat{\alpha}_2\right)}\right),$$

I obtain

$$\widehat{E(FI_3|\text{pa})} = \frac{\sum_{i \in S_1 \setminus (S_2 \cup TA)} \pi(x_{1i})}{\#S_1 - (\#S_2 + \#TA)},$$

where TA is the set of individuals that temporarily non-responded during HUNT2 – these are ignored in the analysis.

Finally, I collect terms:

$$\widehat{E(FI_3)} = \widehat{E(FI_3|R_2=1)} \times \frac{\#S_2}{\#S_1 - \#TA} + \widehat{E(FI_3|\text{pa})} \times \frac{\#S_1 - (\#S_2 + \#TA)}{\#S_1 - \#TA}, \tag{B.15}$$

where $\widehat{E(FI_3|R_2=1)}$ can be computed as either (B.11) or (B.14). The figures in the main text are based on $\widehat{E(FI_3|R_2=1)}$ computed via 'imputation,' but calculations available on request show that $\widehat{E(FI_3|R_2=1)}$ computed via inverse-probability weighting produces similar results. Thus up the $\#TA$ temporary attritors, (B.15) corrects for non-response.

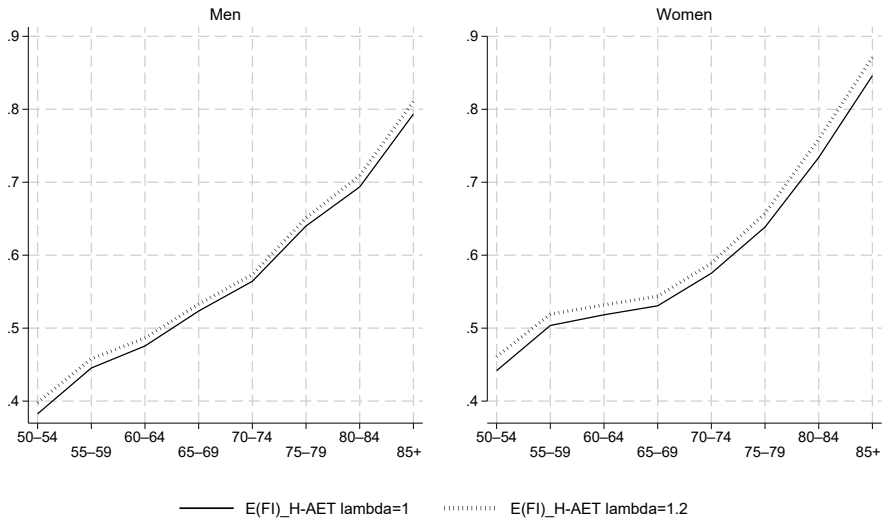
B.5 Sensitivity analysis: 'SoU=1.2×SoO'

In the main text, I presented the age profiles for the prevalence rate of FI, $FILE$, and its complement, $DFLE$, under the assumption that selection on unobservables is as strong as selection on observables. I also gave an intuitive argument for why the latter degree of selection is likely to be a lower bound for selection on unobservables, and that we may expect this type of selection to be stronger if recent health information is moved from the vector of observables into the error term. This intuition was formalised in Section B.1. Because the model for HUNT3 is estimated on $x_3 = (x_1, FI_2)$, some of the health information that is relevant and recorded in HUNT2 (SAH, BMI, incidence of diseases) is forced to be treated as unobservable because it is not recorded for non-participants in HUNT2. Therefore it may be expected that for HUNT3 'SoU $>$ SoO,' and one way of

finding out how this affects results is to estimate the model on HUNT3 data under the assumption that $\rho_3 = 1.2 \times \frac{cov(x_3'\alpha_3, x_3'\beta_3)}{var(x_3'\beta_3)}$. The maximum likelihood estimation results are available upon request. In a nutshell, (1) most covariates get coefficients that are close to their estimates under $\rho_3 = \frac{cov(x_3'\alpha_3, x_3'\beta_3)}{var(x_3'\beta_3)}$, (2) the estimated value for ρ_3 changes from -0.350 to -0.433 for men and from -0.536 to -0.671 for women, and (3) the intercept in the FI^* equation increases by about 0.2 in both cases, raising the level of the FI prevalence rates for both genders.

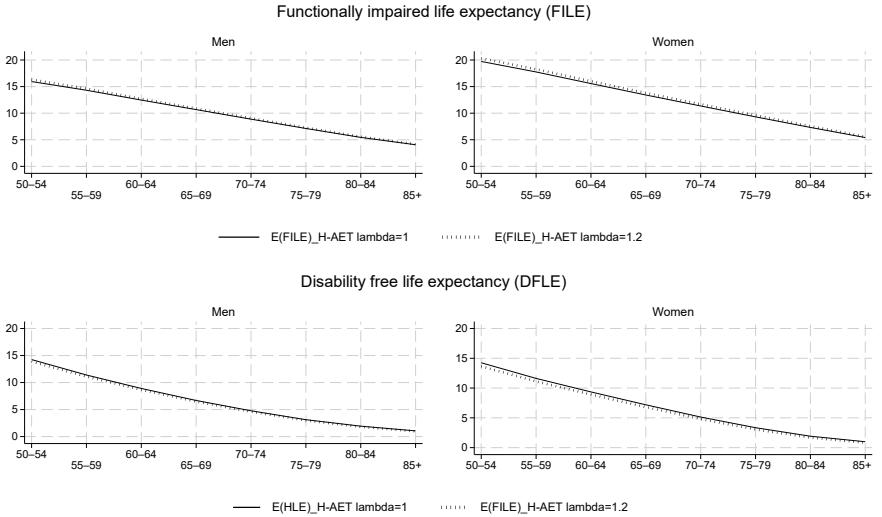
In Figures B.5-1, B.5-2, and B.5-3, the new age profiles for FI prevalence, $FILE$, and $DFLE$ and changes in $FILE$ and $DFLE$ are displayed with dots, whereas the reference profiles (under the 'SoU=SoO' assumption) are drawn with solid lines.

Figure B.5-1: Age profiles for the prevalence rate of FI during HUNT3 when the relative degree of selection on unobservables is raised from 100% to 120%



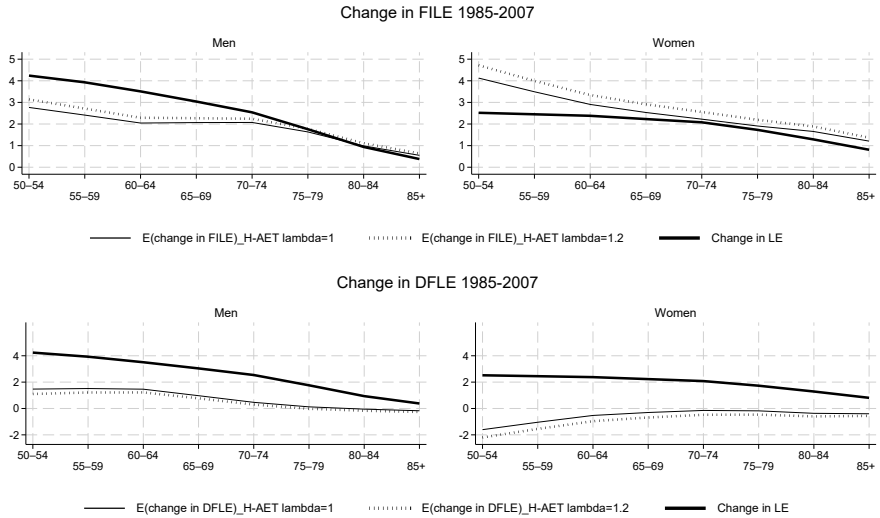
Note: $E(FI)_{\text{H-AET},\lambda=1}$ is calculated as explained in Section 4.2 using maximum likelihood estimates of (13) under the restriction that $\rho_3 = \frac{cov(x_3'\alpha, x_3'\beta)}{var(x_3'\beta)}$. $E(FI)_{\text{H-AET }\lambda=1.2}$ is calculated in the same way, but based on maximum likelihood estimation of (13) under the restriction that $\rho_3 = 1.2 \times \frac{cov(x_3'\alpha, x_3'\beta)}{var(x_3'\beta)}$.

Figure B.5-2: Age profiles for $FILE$ and $DFLE$ during HUNT3 when the relative degree of selection on unobservables is raised from 100% to 120%



Note: $E(FILE)_{\text{H}-\text{AET } \lambda=1}$ is calculated using the Sullivan method, based on the HUNT3 age profile for the estimated FI prevalence $E(FI)_{\text{H}-\text{AET } \lambda=1}$. $E(FILE)_{\text{H}-\text{AET } \lambda=1.2}$ is computed in the same way, but now based on the HUNT3 age profile for the estimated FI prevalence $E(FI)_{\text{H}-\text{AET } \lambda=1.2}$.

Figure B.5-3: Age profiles for the change in $FILE$, $DFLE$, and LE (1985–2007) when the relative degree of selection on unobservables is raised from 100% to 120%



Note: E(change in $FILE)_{\text{H}-\text{AET }\lambda=1}$ is the difference between $E(FI)_{\text{H}-\text{AET }\lambda=1}$ for HUNT3 and $E(FI)_{\text{raw}}$ for HUNT1. E(change in $FILE)_{\text{H}-\text{AET }\lambda=1.2}$ is the difference between $E(FI)_{\text{H}-\text{AET }\lambda=1.2}$ for HUNT3 and $E(FI)_{\text{raw}}$ for HUNT1. Change in LE is the difference between life expectancy in 2007 and 1985. Change in $DFLE$ equals change in LE minus change in $FILE$.