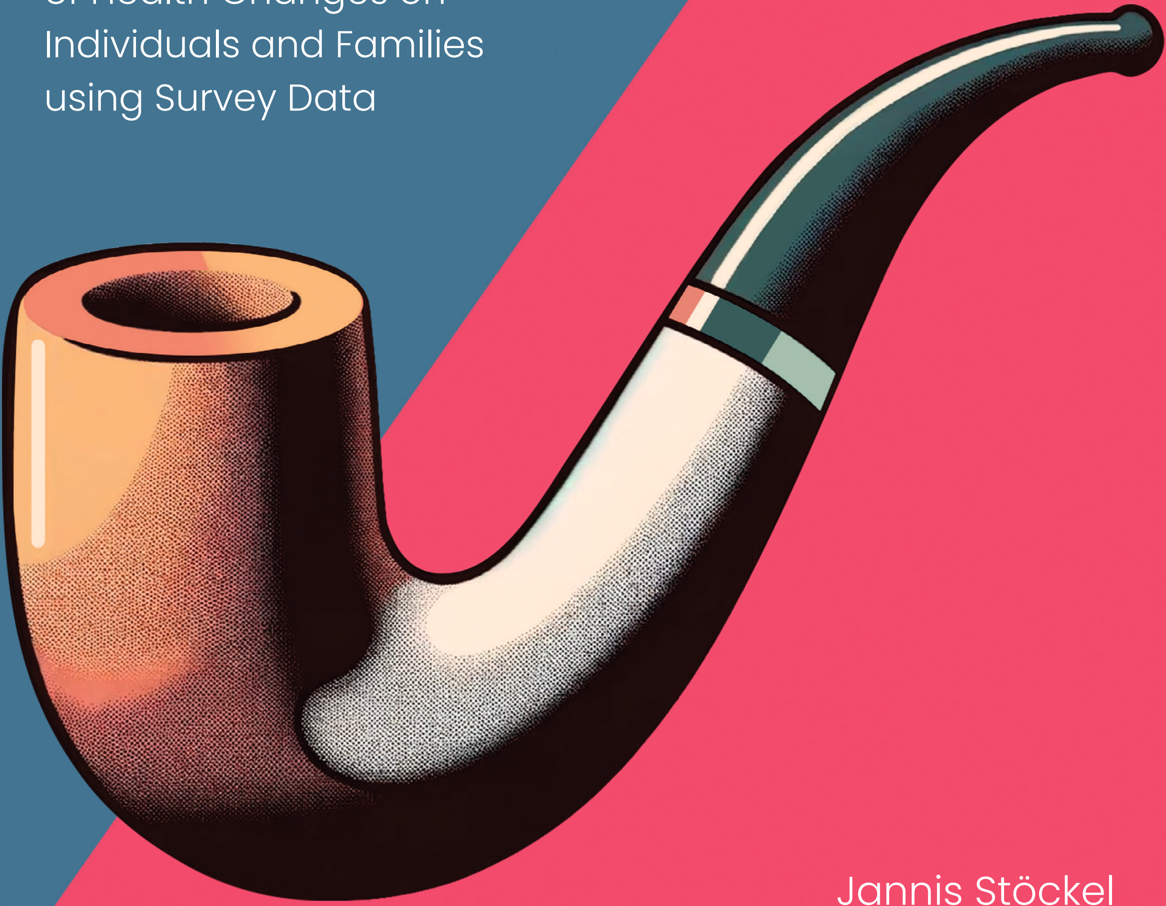
SUBJECTIVE PERSPECTIVES (F)OR OBJECTIVE TRUTHS?

Quantifying the Impact of Health Changes on Individuals and Families using Survey Data



Jannis Stöckel

Subjective Perspectives (f)or Objective Truths?

Quantifying the Impact of Health Changes on Individuals and Families using Survey Data

Jannis Stöckel

Colophon

ISBN: 978-94-6483-647-9
Printing: Ridderprint, www.ridderprint.nl
Layout and design: Anna Bleeker, persoonlijkproefschrift.nl

Subjective Perspectives (f)or Objective Truths?
Quantifying the impact of health changes on individuals and families using survey data

Subjectieve perspectieven of objectieve waarheden?
Het kwantificeren van het effect van gezondheidsveranderingen op individuen en families met behulp van vragenlijstgegevens

Thesis

to obtain the doctoral degree at the
Erasmus University Rotterdam
upon the authority of the
Rector Magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Friday 19 January 2024 at 13:00hrs
by

Jannis Stöckel
born in Quakenbrück, Germany

Erasmus University Rotterdam

Doctoral Committee:

Promotor(s): Prof.dr. N.J.A. van Exel
Prof.dr. W.B.F. Brouwer
Prof.dr. E.W. de Bekker-Grob

Other members: Prof.dr. M.G. Knoef
Prof.dr. O.A. O'Donnell
Prof.dr. V. Angelini

TABLE OF CONTENTS

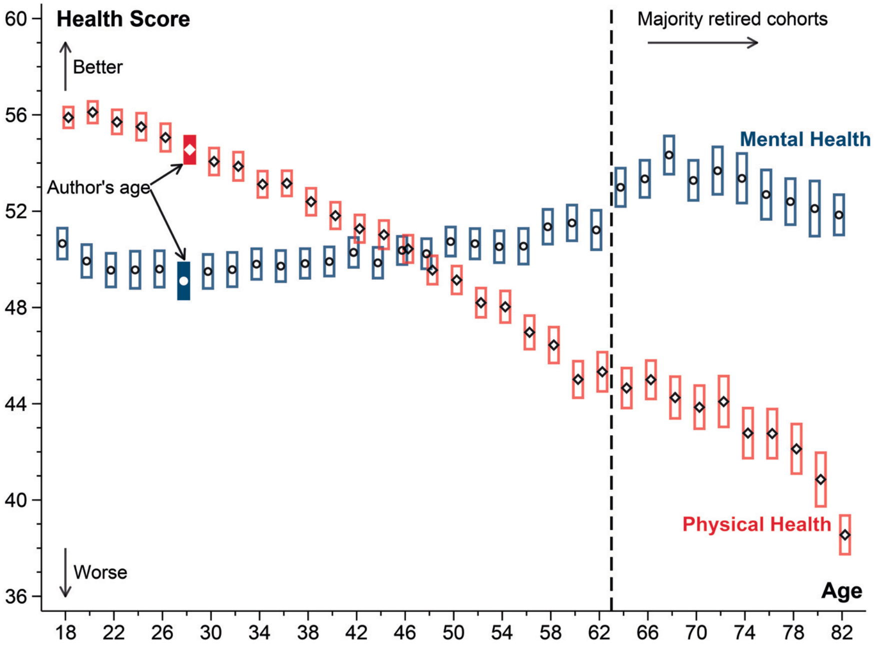
1

CHAPTER 1

Thesis Rationale

The author of this dissertation was born in Germany, a high-income Western European country, and is aged 29 at the time of writing this sentence. If you were to predict his mental and physical health for the coming decades, conditional on survival to a certain age and ignoring the potential impact of future medical innovations, birth-cohort-effects, or other factors like life-style habits or pre-existing medical conditions, a starting point could be Figure 1, depicted below. It depicts the average physical and mental health score by age for a representative cross-section of the German population, based on a multi-dimensional survey-based health measure collected in 2018.

Figure 1: Mental and Physical Health over the Life-Course



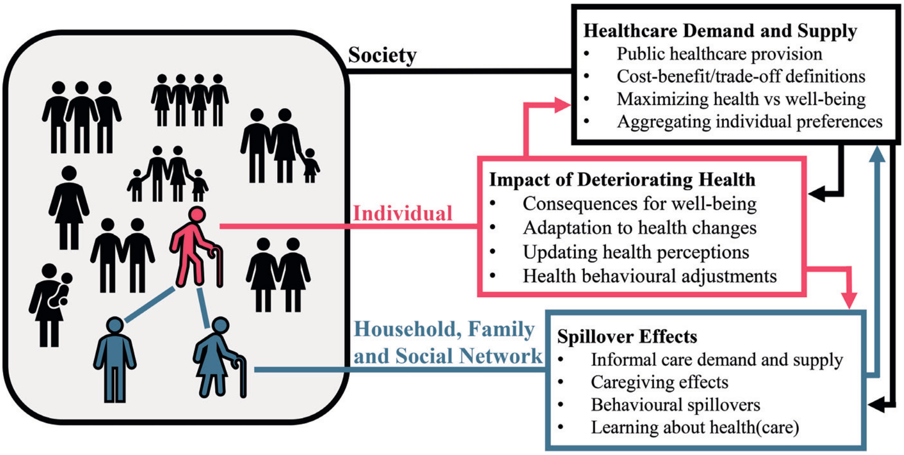
Source: Own illustration using the German Socio-Economic Panel, wave 35 (2018). *Note:* The short-form 12-item health questionnaire physical and mental health scores range from 0 (worst) to 100 (best) with a mean of 50 and standard deviation of 10. Bars indicate 95% confidence intervals.

The main message seems bleak. Starting already in the mid-20s, physical decline sets in for the average German. For the author this implies that he might already be on a downward trajectory, having (unwittingly) passed his prime in terms of physical health even before making any keystroke in the writing of this dissertation. On the upside, however, his mental health might be expected to improve gradually in the coming years (in his special case possibly in relation to completing a dissertation), although larger gains could remain wanting until retirement edges closer.

The purpose of this figure is not to underline a depressing reality about the author's expected health trajectory. At its core it makes a fundamentally true and thereby trivial point: Physical health tends to decline over the human lifespan. However, it does not require an interest in the economics of health sufficient to result in doctoral studies to recognise that what holds true at the population level does not necessarily translate into relevant insights for a given individual. Health trajectories are complex. Even within highly localised populations a multitude of patterns can be observed with distinct profiles of physical health decline over the life-course that are only partially attributable to the usual suspects such as lifestyle habits or socio-economic circumstances (see e.g., Dieteren et al., 2020). Or phrased in less abstract terms; not everyone's physical health heads downwards in a similar fashion -as portrayed in Figure 1- and not always for the same, if any, identifiable reasons. The point is, however, that eventually, whether due to a gradual decline or a sudden shock, everyone is inevitably confronted with health declines in some shape or form. And even if such declines are not experienced personally but observed in friends or family-members, they underline the importance of health as an economic resource and the many consequences that come with its decline.

Identifying and quantifying these consequences of health declines is the core motivation for the chapters included in this dissertation. However, just as individual-level health trajectories are complex, so do the consequences of declining health vary depending on the perspective taken. This dissertation explores three distinct topics illustrating the diverse impacts of health declines at the individual level, the spillover effects into the social network of those experiencing health declines (e.g., households and families), and the broader societal perspective in publicly funded healthcare systems, designed to meet the healthcare demands associated with individual-level health changes. The three explored impact dimensions and the related topics explored alongside their connections are outlined in Figure 2.

Figure 2: Impact Dimensions of Health Declines



Source: Own illustration based on this dissertation. *Note:* Vertical order of dimensions follows the order of discussion within the parts and chapters of this dissertation.

Choices on how to allocate public resources in the provision of healthcare are particularly uncomfortable as often all alternative spending options seem deserving if they improve health outcomes one way or the other. Resolving such trade-offs requires an understanding of how much society values the health gains associated with different choice-options compared to their (monetary) costs. The individual-level well-being consequences experienced due to declining health can help inform these societal trade-offs when translated into monetary equivalents. *Valuing Health Changes using Survey Data* on well-being outcomes for the purpose of informing decision making on the allocation of healthcare resources is the first topic of this dissertation. The second topic focuses on the dimension of individual-level consequences of declining health by exploring *The Impact of Health Changes on Perceptions and Behaviours* through the lens of survey-based subjective outcomes measures. It first explores whether experienced losses in well-being and subjective health following the onset of physical disability diminish over time due to adaptation. Subsequently it considers the role of severe health events as sources of information about one's own health by exploring how heart attacks and strokes shape the health perceptions and behaviours of patients over time. Lastly this dissertation turns towards the *Spillovers Effects of Declining Health* which is the third topic of this dissertation. Given the important role of within- and between-household support in mitigating the consequences of health shocks through informal care provision, this part considers spillover effects on family-caregivers' own health and how health shocks shape health perceptions and behaviours within social networks. The three topics of this dissertation are interconnected by a common driver, that is, individual level health changes over the life course. They are also connected by their focus on survey-based subjective outcome measures as an instrument to quantify the multi-dimensional impacts of health declines.

Topic 1: Valuing Health Changes using Survey Data

In most industrialized economies, gains in life-expectancy and decreasing fertility rates drive an increasing demand for medical care with a simultaneous shrinking of the working population. Already now, members of the Organisation for Economic Co-operation and Development (OECD), spend one tenth of their gross domestic product annually on healthcare, with projected expenditures only increasing (OECD, 2019). These trends pose a challenge to healthcare systems aiming to remain financially sustainable. With only few exceptions OECD member states already rely heavily on public entities within their healthcare systems to directly deliver healthcare or use public resources to fund or subsidize private healthcare providers. Public resources are, however, not endless, with many policy domains such as education, environmental protection, or national defence competing in their allocation. Spending on any healthcare intervention, therefore, has opportunity costs represented in spending alternatives within the healthcare sector but also in other domains, highlighting the need to efficiently allocate resources (Hendren & Sprung-Keyser, 2020). Ultimately, this implies trade-offs and uncomfortable choices: How much health can be produced for whom and at what price? And are some medical interventions or public health policies simply too costly, given existing spending alternatives with equal or greater benefit within the healthcare sector or elsewhere?

Economists resolve such trade-offs using welfare-economic tools like cost-benefit and cost-utility analyses. In the healthcare context, however, calculating, comparing, and aggregating all costs and benefits associated with choice alternatives is not always straightforward. A first question is what the maximand of a given policy and, therefore, the target of any cost-benefit exercise should be. The benefits of implementing a given public health intervention that are to be maximized while keeping costs low might seem obvious, but the exact definition based on which benefits are measured is not. Heart attack incidence, 5-year cancer survival rates, life expectancy at birth, self-reported health status, and overall quality of life or life satisfaction are just some examples of highly specific or broadly defined outcomes that could capture the gains to be maximised by (health) policy. Choosing one, or some, of these possible outcomes determines the evaluative scope of what constitutes the benefits derived from improving certain dimensions of population health as health is an important input to individual-level well-being, but not its sole determinant (Sen, 1993). Recognizing this alongside the role of spillovers on other groups like caregivers or the heterogeneities existing across patient populations with respect to how certain health conditions impact individual level well-being, can provide a drastically different picture than a focus on physical health changes only (Brouwer et al., 2008).

These considerations feed into the practical application of decision-making rules in the economic evaluation of healthcare policies and new technologies. Health policy interventions aimed at the early stages of the human lifespan, i.e., new-borns and children, often produce large benefits across multiple dimensions of (economic) well-being and health at a comparatively small initial investment (Hendren & Sprung-Keyser, 2020), making them highly cost-effective. Policies aiming at later stages of life, on the other hand, have a reduced timeframe in which benefits can be realized and aggregated, while physical health benefits might be inherently limited for older and/or less healthy populations. Cost-effectiveness when evaluated purely based on physical health gains could be hard to achieve given these population-specific constraints. However, the same intervention might be evaluated differently when considering gains in broader well-being as measured in terms of self-reported life satisfaction or other subjective quality of life measures. This underlines a core dilemma in the allocation of healthcare resources; how much is too much, the threshold at which the cost per unit of outcome produced is too high given alternatives for public spending? And how should these values be calculated and compared across alternatives and populations of potential beneficiaries?

There is a range of methods available to identify such thresholds. A "supply-side" perspective could rely on the production of health using existing healthcare technologies and costs or the consumption value of certain health or income gains (Claxton et al., 2011) to identify spending options creating equivalent benefits. Alternatively, given the use of public resources, the broader preferences of members of the public can be used to obtain willingness-to-pay estimates based on stated preference exercises that directly elicit "demand-side" estimates on how much people are willing to pay for health improvements (Ryen & Svensson, 2015). Each of these approaches come with their own normative assumptions and methodological limitations. For example, by asking for an evaluation of hypothetical health states, attention can be drawn towards certain aspects of (physical) health like eyesight or specific forms of physical functioning. This is often a deliberate methodological decision in a study but also a potential source of bias alongside the fact that many experimental participants will rely on imagined but not first-hand experiences in the evaluation of health states and their impact on day-to-day lives (Dolan & Kahneman, 2008). On the one hand, this is a choice consistent with the idea that the broader public's preferences should influence allocative decision making, but it does result in ethical dilemmas if ex-ante (imagined) and ex-post (experienced) preferences differ (Brazier et al., 2018). Further, the resulting valuations based on such methods often depict a high sensitivity towards experimental set-ups e.g., the framing of evaluation tasks, that make it hard to compare estimates across studies and contexts (Ryen & Svensson, 2015).

In the context of determining the monetary value of health, survey-based subjective outcome measures on broader well-being provide a complementary perspective. Collected at the scale of representative general population samples they allow to observe the experienced impact of health changes on subjective well-being (Dolan & Kahneman, 2008). While some economists tend to discount survey-based measures on subjective well-being outcomes (Bond and Lang, 2019) they are consistent predictors of economic variables and individual-level behaviours across broad dimensions, therefore presenting an informative proxy for revealed preferences or experienced utility (Kaiser & Oswald, 2022). By exploiting the variation of income, health, and well-being changes over time at the individual-level this allows for an estimation of the monetary equivalent of health changes using methods such as the well-being valuation approach that estimates the compensating income necessary to offset well-being losses due to deteriorating health. As large-scale general population (longitudinal) surveys are widely available across countries, the application of such methods could provide an alternative set of estimates on the monetary value of health that are easier to compare than those based on stated-preference studies (Huang et al., 2018). However, the reliance on econometric methods as opposed to experimental designs to identify the parameters of interest also comes with distinct challenges that need to be understood to evaluate the strengths and limitations of the well-being valuation approach in this context.

Topic 2: The Impact of Health Changes on Perceptions and Behaviours

Adaptation in Subjective Well-being and Health

Survey-based subjective outcome measures have the advantage of capturing the perceived consequences of declining health and the resulting well-being changes (Dolan & Kahneman, 2008). A caveat of relying on measures of individual experiences associated with declining health is how to account for heterogeneity across the population in the experience of ill health and to assure the intertemporal validity of such measures given that health is only one of many well-being determinants over the life-course. One manifestation of intertemporal comparability being violated is the phenomenon of adaptation. In the context of changing health states, adaptation means that in subjective outcomes at the onset of a medical condition or disability are not indicative of the longer-term impact as individuals adapt to their new health state (Frederick & Loewenstein, 1999). The mechanisms that could underlie adaptation are varied.

Individuals might adjust their daily lives using medical devices or other services that alleviate the burden of disease, or substitute for reduced physical functionalities. On the other hand, adaptation could also be driven by psychological resilience and coping mechanisms (Etilé et al., 2021). This form of adaptation might manifest itself in ignoring or denying the impact of ill health or by finding new sources of well-being that compensate those lost because of a change in physical functioning. Lastly, transitioning into a state of worse health could itself alter how individuals evaluate dimensions of health and well-being. It might shift the reference point or comparison group applied and, thereby, alter the scale on which different aspects of health (e.g., a specific physical function) are evaluated in the light of decreased expectations of future health (Lindeboom & van Doorslaer, 2004; Loewenstein & Ubel, 2008).

When survey-based subjective outcome measures such as self-perceived health or well-being are used to quantify the impact of health changes, adaptation is a concern irrespective of which mechanism is driving it (Groot, 2000). It raises a range of practical and normative concerns on how adaptation should be accounted for when it occurs. Absence of adaptation can be seen as an indicator of the true burden of disease as experienced by individuals. There is a well-documented heterogeneity with respect to how different chronic physical conditions impact well-being (see e.g., Howley, 2017) and with respect to the differential impact of mental versus physical health changes (see e.g., Graham et al., 2010). Adaptation predominantly occurring in the case of physical health changes can be seen as an argument for preferring broader well-being-based measures as they depict this true impact of certain health conditions like those impacting mental health (Peasgood, Foster, and Dolan, 2019). On the other hand, while survey-based subjective outcome measures have been shown to predict decisions (Caspar & Oswald, 2022) individuals themselves envisioning the impact of certain life events on their subjective outcomes do not account for their own adaptation (Odermatt & Stutzer, 2019). In other policy contexts in which behavioural economic reasoning is applied, such misalignments would be seen as a case where policymakers can improve welfare by addressing these discrepancies (Dolan & Kahnemann, 2008). Doing so does, however, have serious implications if the propensity to adapt is not equally distributed across the population but is a matter of psychological and other coping resources (Etilé et al., 2021). This puts allocative decisions that assume adaptation to occur as a universal phenomenon at risk of disadvantaging certain patient groups (Cohen 1993; Menzel et al., 2002). This underlines the need to consider the temporal dimension and the potential for adaptation when relying on survey-based subjective outcome measures to quantify the impact of individual-level health changes.

The Impact of Health Shocks on Health Perceptions and Behaviours

Survey-based subjective outcome measures also allow to understand how the experience of health changes shapes the subjective assessment of individuals' own health status. These health perceptions play an integral role in theoretical models of health-related decision-making. The classic economic model of individual-level health production centres around the individual's decision to make costly investments counter-acting deteriorating health (Grossman, 1972). A fundamental assumption underlying this is that health is observed and assessed accurately by the individual decision maker. However, outside the well-behaved world of economic models in which decision makers possess perfect information based on which they form perceptions about their own health status, the role of objective health information is less clear. As health changes are inherently uncertain (Cropper, 1977), it is difficult to directly infer from having a specific condition, like hypertension, how healthy one is given that such a condition can impact both current health but also the risk to suffer from future health shocks such as a heart attack. Also, the ability to accurately process and understand health-related information and the potential consequences of certain risky health habits, like smoking, or excessive alcohol consumption, is a pre-requisite for optimal decision making that is not guaranteed. Engaging in risky health behaviours has been found to be highly correlated with education and general cognitive abilities, in part explaining the well-documented education health gradient next to other important factors such as income differences and healthcare access (Cutler & Lleras-Muney, 2010).

Based on empirical evidence, however, there is a reliable link between objective health information and subjective health perceptions in most circumstances. Across populations, measures of objective health like chronic conditions, hospital visits or functional limitations, are highly correlated with broader measures of perceived health status (Jürges, 2007). The informational basis of self-perceived health is not only confined to current health outcomes. Individuals seem to incorporate various sources of information amalgamating past health outcomes as well as the implications of current health behaviours for future health prospects (Nielsen, 2016). At the same time there is considerable heterogeneity in how objective health outcomes precisely map onto health perceptions with potential implications for individual-level decision making. While across populations there is a consistent link between objective and subjective health, the strength of this relationship and the importance of specific objective health dimensions vary considerably across countries (Jürges, 2007) and age-groups (Au & Johnston, 2014). In addition, the same behavioural biases found in other domains, like optimism and pessimism biases, are also found with respect to how some individuals perceive their own health to be (see e.g., Bago d'Uva et al., 2020). These biased perceptions are likely to not just be an instance of survey-based outcome measures being an unreliable proxy with no bearing on decision making, as they are associated with an increased engagement in risky health behaviours (Arni et al., 2021) and lower usage of preventive healthcare services (Spitzer and Shaikh, 2022).

In this context, health shocks can be seen as sources of information based on which individuals might update their health perceptions using new information on their objective current and future health and the consequences of past behaviours. However, as discussed in the context of adaptation, the degree to which changes in self-perceived health are persistent over time is not clear. A range of studies have shown that adaptation in broader well-being to disability is in part driven by adaptation in self-perceived health (see e.g., Powdthavee 2009; de Hondt et al., 2019). Similar patterns have been observed in the case of severe health shocks that seem to only have a transient effect on health perceptions and, therefore, insufficiently capture longer-term changes to objective health trajectories and survival probabilities (Baji & Biro, 2018). Such findings could have serious implications for observed behavioural patterns and help explain the common finding of highly persistent health habits (see e.g., Royer et al., 2015) and low adherence to long-term medication therapies (Choudhry et al., 2011). Exploring how health shocks shape health perceptions and behaviours could yield important insights on the determinants of individual-level health beliefs and behaviours over the life course. However, doing so is empirically challenging as it requires a clean identification of the impact of health events on both dimensions to explore how these changes relate to objective health changes and pre-existing health information individuals had. To explore whether health events do result only in short-term updating of health perceptions or persistent learning effects and adjustments to individual-level health behaviours, therefore, requires the combination of information on survey-based subjective outcome measures and detailed information on individual-level objective health trajectories.

Topic 3: Spillover Effects of Declining Health

Health Shocks and Informal Care in the Household and Family Network

Beyond their impact on individuals whose health is directly affected, health shocks can create spillovers within households and the wider family or the social network of patients. If family members intrinsically care about the health and well-being of those within their network, a health shock to one might result in direct negative mental health effects to others in their network (Bobinac et al., 2010). Beyond this, health shocks also have the potential to alter the demand and supply of support provided by the members of the household or broader family network. If the individual affected ceases to work this might require others to take up additional paid employment to compensate for lost income (Lundberg, 1985). At the same time, health shocks can cause competing demands. Temporary or permanent disability following health shocks might require support in daily activities and the involvement of family members, friends and most often partners, in managing personal tasks and the delivery of care. This support is commonly referred to under the umbrella term of unpaid (family) care, or also informal care; the provision of health and non-health related personal care tasks to individuals in need within one's social network.

Across developed economies informal care produced between and within households is already a key pillar in meeting the demand for long-term care for ageing populations and supporting ageing-in-place policies (Colombo et al., 2011). In this context informal care can have distinct advantages. Among care recipients as well as informal caregivers, it might be preferred over formal care as it avoids institutionalisation or and increases social contacts between both. From the perspective of public healthcare providers, it helps delaying usage of high-cost nursing home care or shortening hospitalisations without evidence for trading-off care-quality (van Houtven & Norton, 2004; 2008; Barnay & Juin, 2016; Coe et al., 2019). It is also often low cost with respect to direct costs incurred by the public and private households (e.g., in terms of associated labour unit costs), as monetary support to informal caregivers is only a fraction of equivalent formal care labour-costs (Courtin et al, 2014). In some circumstances, and focusing purely on the care-recipient, weighting direct costs associated with supplying informal care against the generated benefits makes a compelling case for governments to incentivise reliance on family care. Whether this is an efficient policy choice from the societal perspective, however, depends on whether informal care produces relevant spillovers on caregivers themselves and how and what type of spillovers are considered in its evaluation.

By decreasing working hours or time for leisure activities or directly impacting mental and physical health due to strenuous or stressful caregiving tasks, informal caregiving can affect the economic well-being and health of caregivers. A purely fiscal perspective might weigh the benefit of low-cost care produced by households against the resulting economic costs such as reduced labour market productivity and healthcare costs generated by caregivers' care consumption. For both there is little evidence that these costs are sufficiently large to offset the associated gains (Bauer and Sousa-Poza, 2015) with labour market outcomes protected by comprehensive public social insurance schemes prevalent in most European countries (Rellstab et al., 2020; Fadlon & Nielsen, 2021). Such a narrow definition, however, omits spillover effects in domains which are hard to quantify based on market prices, such as lost time for leisure activities or changes in overall well-being and mental health not directly resulting in increased healthcare demand (van den Berg & Ferrer-i-Carbonell, 2007). This requires a better understanding of how informal care provision affects the health and broader well-being of caregivers with respect to how the caregiving context, such as the broader policy environment, the caregiver-recipient relationship or caregiving intensity and duration shape spillovers. Survey-based subjective outcome measures provide a crucial lens for identifying such spillovers in domains that are relevant to the societal impact of informal caregiving, such as mental health and broader well-being.

Health Shocks and Spouses' Perceptions and Behaviours

Spillovers of health shocks between and within households are not confined to the household- or family-network's production of informal care but could directly influence members' individual decision-making. Even if individuals are not experiencing a shock to their own health, they might react to such an event occurring in their direct family network given its informational nature. Spouses, but also household and other family members more broadly, depict a high level of concordance in their general health status (Wilson, 2002; Banks et al., 2021), engagement in risky health behaviours (Meyler et al., 2007; Cutler & Lleras-Muney, 2010; Cawley & Ruhm, 2011) and health prevention efforts (Bouckaert et al., 2020). Observed health shocks connected to shared (behavioural) risk factors (e.g., smoking or alcohol consumption) or pre-existing conditions (e.g., hypertension or diabetes) are informative also to those individuals not directly affected. Even in absence of known shared risk factors, an observed health shock can be insightful. Observing the consequences of a severe deterioration of health can function as a salient cue highlighting the need for health insurance or other protective behaviours and generally prompting individuals to act (Bernheim & Rangel, 2004). The subsequent contact with the healthcare sector might also provide new information about their own health or health-related information more broadly on risk factors or how to navigate the healthcare system (Chen et al., 2022). Individual-level health shocks could therefore propagate their effects beyond the directly impacted person onto health perceptions and behaviours of social network members.

A growing literature has documented this causal link between health shocks and broader behavioural responses among social network members. Severe health shocks, such as heart attacks or strokes, lead to an increase in preventive health behaviours among spouses, adult children, and even co-workers (Fadlon & Nielsen, 2019). There are two potential mechanisms underlying this propagation of spillover effects at the level of the individual decision-maker: learning new information and salience, the subjectively perceived importance or relevance of an event (e.g., Bernheim & Rangel, 2004; DellaVigna, 2009; Bordalo et al., 2013). These two phenomena are also connected. Without a minimum level of salience even high-information events might not result in learning or behaviour changes. Likewise, salience might also draw attention to specific events or aspects of an event which are not the most informative given individuals' personal health risk profile. Particularly severe health shocks might therefore draw attention towards highly specific but unlikely events with behavioural responses characterized more by emotional drivers or an amplified perception of how likely such events are than informed re-evaluation of risk factors based on newly acquired information.

In the case of family spillovers in health behaviours, the relative importance of these mechanisms remains an open question. Overall, there is a high domain-specific nature of documented behavioural responses indicating salience as a key mechanism. Following cardiovascular health shocks, preventive efforts specific to cardio-vascular health increase, even among those individuals that received information on this health dimension shortly before the shock occurred (Fadlon and Nielsen, 2019). However, there is also evidence that spouses and other family members react more broadly to observed family health shocks in areas beyond the health shock's domain, possibly learning about more about health risks, their insurance coverage or healthcare access (Hodor, 2021). Understanding the underlying behavioural mechanisms in the response to family health shocks is crucial to design effective public health policies and information campaigns. Whether observed family health shocks change individuals' perceptions of their current health, the perceived risk of certain events, or drive emotional responses through their high salience, imply different lessons for effective policy design, balancing the use of the window of opportunity following health events against discouraging inefficient care consumption (Hoagland, 2021). In this context survey data on subjective health perceptions in combination with data on behavioural habits and care consumption could help illuminate the relative importance of these mechanisms and answer the question whether family-members own health perceptions are altered by the health event observed within their family-network. Understanding whether family spillovers are a commonly occurring phenomenon and what behavioural mechanisms underlie them is particularly relevant as it could help inform the design of effective public health interventions.

Objectives and Outline

The chapters of this dissertation explore the multi-dimensional impact of declining health and health shocks on patients and their social network, and how survey-based subjective outcome measures can help quantify these impacts to inform healthcare decision-making at individual and societal level. As outlined above, this dissertation narrows down on three specific topics: (i) quantifying the well-being effects of declining health for the purpose of calculating monetary equivalents of health changes, (ii) the impact of declining health on survey-based subjective outcome measures and health behaviours over time in those experiencing a health shock and (iii) spillover effects on family members, in particular the impact of spousal health shocks on individual-level decision making and the health impact of providing informal care on caregivers themselves. In exploring these topics, the following research questions are answered:

1) How can large scale observational survey data on subjective well-being be used to estimate the monetary value of health?
2) How does the impact of declining health on individual-level well-being evolve over time and across subgroups?
3) How does experiencing severe health shocks change health perceptions and health-related behaviours?
4) Do health shocks result in changes to self-perceived health and risky or preventive health behaviours also in family members not directly experiencing them?
5) What are the short and long-term health effects of providing informal care to family members experiencing ill health?

Chapter 2 explores how individual level experiences of declining health can be leveraged to obtain estimates of the monetary value of health at the population level. Using individual-level survey data from Germany on life satisfaction, income, and health over time it provides an answer to the first research question. It applies the well-being method and considers under which assumptions general population survey data can provide a broader basis for a societal valuation of health changes in monetary terms, and how these estimates relate to existing ones using methodologically different approaches. This requires the estimation of the causal effect of health and income on well-being to calculate the monetary value necessary to compensate individuals for the well-being changes associated with health losses. As such empirical estimations require a range of assumptions, Chapter 2 explores the sensitivity of results to methodological choices relevant for the application of this method in the broader context of health economic evaluations and the informational value of the obtained results.

While Chapter 2 utilizes individual-level experiences to estimate a population-level parameter, Chapters 3 and 4 turn towards the individual-level experiences of deteriorating health through the lens of subjective well-being and self-perceived health. Using survey data from the United Kingdom (UK), Chapter 3 explores the degree to which well-being and self-assessed health change with the onset of temporary or longer-term disability; does life-satisfaction and self-perceived health remain stable after the onset of disability or do individuals adapt to their declined health over time? It further explores the degree to which adaptation is moderated by individual-level characteristics, like age or gender, and the severity of experienced health states. Chapter 3, therefore, answers the second research question, while also providing some evidence on the third research question regarding the impact of health shocks on individual-level health perceptions, which is the focus of Chapter 4. This chapter uses a combination of survey and administrative data from the Netherlands to explore how sudden and severe health shocks like heart attacks and strokes shape self-perceived health, objective health, and health related behaviours of patients over time. To do so, it exploits the onset of unexpected health shocks to estimate the causal effect of such events on health perceptions and how these changes in subjective health relate to objective health changes. Chapter 4 further explores whether such health shocks impact risky health behaviours and the use of preventive medication, which would be consistent with changes in individual-level health perceptions.

Research questions four and five on the impact of health shocks on family members' perceptions of their own health and spillover effects on informal caregivers' health are explored in Chapters 4 to 6. The second half of Chapter 4 considers how spouses react to severe health shocks to their partners (i.e., heart attacks or strokes), documenting their impact on self-perceived health and risky health behaviours and, therefore, answering research question four. Chapters 5 and 6 deal with spillover effects arising from increased healthcare needs in the family due to declining health of parents and spouses and the informal care provided by family members. Using longitudinal survey data from the UK and the Netherlands, it explores the causal impact of providing informal care on caregivers' mental and physical health outcomes. Chapter 5 does so by comparing health effects across the different policy contexts of a highly generous long-term care system (the Netherlands) with a less generous system in which informal care is the main source of social care services (UK). Chapter 6, on the other hand, focuses on the UK but explores the role of the caregiving context and more particularly on duration in explaining the magnitude and persistence of spillover effects on caregivers' mental and physical health.

The final chapter closes this dissertation with a summary of the research findings on each of the three topics before discussing overarching strengths and limitations and the implications of the findings for policy-making and future research directions.

References

Au, N., & Johnston, D. W. (2014). Self-assessed health: what does it mean and what does it hide?. *Social Science & Medicine*, 121, 21-28.

Arni, P., Dragone, D., Goette, L., & Ziebarth, N. R. (2021). Biased health perceptions and risky health behaviors—Theory and evidence. *Journal of Health Economics*, 76, 102425.

Baji, P., & Bíró, A. (2018). Adaptation or recovery after health shocks? Evidence using subjective and objective health measures. *Health Economics*, 27(5), 850-864.

Bago d'Uva, T., O'Donnell, O., & Van Doorslaer, E. (2020). Who can predict their own demise? Heterogeneity in the accuracy and value of longevity expectations. *The Journal of the Economics of Ageing*, 17, 100135.

Banks, J., Kesternich, I., & Smith, J. P. (2021). International differences in interspousal health correlations. *Health Economics*, 30(5), 1152-1177.

Barnay, T., & Juin, S. (2016). Does home care for dependent elderly people improve their mental health?. *Journal of Health Economics*, 45, 149-160.

Bauer, J. M., & Sousa-Poza, A. (2015). Impacts of informal caregiving on caregiver employment, health, and family. *Journal of Population Ageing*, 8(3), 113-145.

Bernheim, B. D., & Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 94(5), 1558-1590.

Bobinac, A., Van Exel, N. J. A., Rutten, F. F., & Brouwer, W. B. (2010). Caring for and caring about: disentangling the caregiver effect and the family effect. *Journal of Health Economics*, 29(4), 549-556.

Bond, T. N., & Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, 127(4), 1629-1640.

Bouckaert, N., Gielen, A. C., & Van Ourti, T. (2020). It runs in the family–Influenza vaccination and spillover effects. *Journal of Health Economics*, 74, 102386.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5), 803-843.

Brazier, J., Rowen, D., Karimi, M., Peasgood, T., Tsuchiya, A., & Ratcliffe, J. (2018). Experience-based utility and own health state valuation for a health state classification system: why and how to do it. *The European Journal of Health Economics*, 19(6), 881--891.

Brouwer, W. B., Culyer, A. J., van Exel, N. J. A., & Rutten, F. F. (2008). Welfarism vs. extra-welfarism. *Journal of Health Economics*, 27(2), 325-338.

Cawley, J., & Ruhm, C. J. (2011). The economics of risky health behaviors. In *Handbook of Health Economics* (Vol. 2, pp. 95-199). Elsevier.

Chen, Y., Persson, P., & Polyakova, M. (2022). The roots of health inequality and the value of intrafamily expertise. *American Economic Journal: Applied Economics*, 14(3), 185-223.

Choudhry, N., Avorn, J., & Glynn, R. (2011). Eliminating Copayments for Patients After Myocardial Infarction Improves Adherence and Some Outcomes. *New England Journal of Medicine*, 365(2011), 2088-97.

Claxton, K., Paulden, M., Gravelle, H., Brouwer, W., & Culyer, A. J. (2011). Discounting and decision making in the economic evaluation of health-care technologies. *Health Economics*, 20(1), 2-15.

Cohen, G. A. (1993). Equality of What? On Welfare, Goods, and Capabilities. *The Quality of Life*, 9-29.

Courtin, E., Jemiai, N., & Mossialos, E. (2014). Mapping support policies for informal carers across the European Union. *HealthPolicy*, 118(1), 84-94.

Cropper, M. L. (1977). Health, investment in health, and occupational choice. *Journal of Political Economy*, 85(6), 1273-1294.

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1), 1-28.

Colombo, F., Llena-Nozal, A., Mercier, J., & Tjadens, F. (2011). Help Wanted? Providing and paying for long-term care. *OECD Health Policy Studies*, OECD Publishing.

de Hond, A., Bakx, P., & Versteegh, M. (2019). Can time heal all wounds? An empirical assessment of adaptation to functional limitations in an older population. *Social Science & Medicine*, 222, 180-187.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2), 315-72.

Dieteren, C. M., Samson, L. D., Schipper, M., van Exel, J., Brouwer, W. B., Verschuren, W. M., & Picavet, H. S. J. (2020). The healthy aging index analyzed over 15 years in the general population: the Doetinchem cohort study. *Preventive Medicine*, 139, 106193.

Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *The Economic Journal*, 118(525), 215-234.

Etilé, F., Frijters, P., Johnston, D. W., & Shields, M. A. (2021). Measuring resilience to major life events. *Journal of Economic Behavior & Organization*, 191, 598-619.

Fadlon, I., & Nielsen, T. H. (2021). Family labor supply responses to severe health shocks: Evidence from Danish administrative records. *American Economic Journal: Applied Economics*, 13(3), 1-30.

Fadlon, I., & Nielsen, T. H. (2019). Family health behaviors. *American Economic Review*, 109(9), 3162-91.

Frederick, S., & Loewenstein, G. (1999). 16 Hedonic adaptation. *Well-Being. The foundations of Hedonic Psychology*/Eds. D. Kahneman, E. Diener, N. Schwarz. NY: Russell Sage, 302-329.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, 19(3), 403-420.

Grossman, M. (1972). The demand for health: a theoretical and empirical investigation. *NBER Occasional Papers*, 119.

Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. The *Quarterly Journal of Economics*, 135(3), 1209-1318.

Hoagland, A. (2021). An ounce of prevention or a pound of cure? The value of health risk information.

Hodor, M. (2021). Family health spillovers: evidence from the RAND health insurance experiment. *Journal of Health Economics*, 79, 102505.

Howley, P. (2017). Less money or better health? Evaluating individual's willingness to make trade-offs using life satisfaction data. *Journal of Economic Behavior & Organization*, 135, 53-65.

Huang, L., Frijters, P., Dalziel, K., & Clarke, P. (2018). Life satisfaction, QALYs, and the monetary value of health. *Social Science & Medicine*, 211, 131-136.

Jürges, H. (2007). True health vs response styles: exploring cross-country differences in self-reported health. *Health Economics*, 16(2), 163-178.

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences*, 119(42), e2210412119.

Lindeboom, M., & Van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 23(6), 1083-1099.

Loewenstein, G., & Ubel, P. A. (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*, 92(8-9), 1795-1810.

Lundberg, S. (1985). The added worker effect. *Journal of Labor Economics*, 3(1, Part 1), 11-37.

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social Science & Medicine*, 55(12), 2149-2158.

Meyler, D., Stimpson, J. P., & Peek, M. K. (2007). Health concordance within couples: a systematic review. *Social Science & Medicine*, 64(11), 2297-2310.

Odermatt, R., & Stutzer, A. (2019). (Mis-) predicted subjective well-being following life events. *Journal of the European Economic Association*, 17(1), 245-283.

OECD (2019), Health at a Glance 2019: OECD Indicators, OECD Publishing, Paris.

Peasgood, T., Foster, D., & Dolan, P. (2019). Priority setting in healthcare through the lens of happiness. *Global Happiness and Well-Being Policy Report*.

Powdthavee, N. (2009). What happens to people before and after disability? Focusing effects, lead effects, and adaptation in different areas of life. *Social Science & Medicine*, 69(12), 1834-1844.

Rellstab, S., Bakx, P., Garcia-Gómez, P., & Van Doorslaer, E. (2020). The kids are alright-labour market effects of unexpected parental hospitalisations in the Netherlands. *Journal of Health Economics*, 69, 102275.

Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, 7(3), 51-84.

Ryen, L., & Svensson, M. (2015). The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Economics*, 24(10), 1289-1301.

Sen, A. (1993). Capability and well-being73. *The Quality of Life*, 30, 1-445.

Spitzer, S., & Shaikh, M. (2022). Health misperception and healthcare utilisation among older Europeans. *The Journal of the Economics of Ageing*, 22, 100383.

Van den Berg, B., & Ferrer-i-Carbonell, A. (2007). Monetary valuation of informal care: the well-being valuation method. *Health Economics*, 16(11), 1227-1244.

Van Houtven, C. H., & Norton, E. C. (2004). Informal care and health care use of older adults. *Journal of Health Economics*, 23(6), 1159-1180.

Van Houtven, C. H., & Norton, E. C. (2008). Informal care and Medicare expenditures: Testing for heterogeneous treatment effects. *Journal of Health Economics*, 27(1), 134-156.

Wilson, S. E. (2002). The health capital of families: an investigation of the inter-spousal correlation in health status. *Social Science & Medicine*, 55(7), 1157-1172

2

CHAPTER 2

The Value of Health – Empirical Issues when Estimating the Monetary Value of a Quality-Adjusted Life Year based on Well-being Data

Based on: Himmler, S., Stöckel, J., van Exel, J., & Brouwer, W. B. (2021). The value of health—Empirical issues when estimating the monetary value of a quality-adjusted life year based on well-being data. *Health Economics*, 30(8), 1849-1870.

Abstract:

Decisions on interventions or policy alternatives affecting health can be informed by economic evaluations, like cost-benefit or cost-utility analyses. In this context, there is a need for valid estimates of the monetary equivalent value of health (gains), which are often expressed in € per quality-adjusted life years (QALYs). Obtaining such estimates remains methodologically challenging, with a recent addition to the health economists' toolbox, which is based on well-being data: The well-being valuation approach. Using general population panel data from Germany, we put this approach to the test by investigating several empirical and conceptual challenges, such as the appropriate functional specification of income utility, the choice of health utility tariffs, or the health state dependence of consumption utility. Depending on specification, the bulk of estimated € per QALY values ranged from €20,000-60,000, with certain specifications leading to more considerable deviations, underlining persistent practical challenges when applying the well-being valuation methodology to health and QALYs. Based on our findings, we formulate recommendations for future research and applications.

1 Introduction

During the ongoing COVID-19 pandemic, many citizens for the first time directly observe scarcity of goods in the health care sector in terms of testing, ventilation, vaccination capacity, and the prioritisation of services under binding capacity constraints. This scarcity and the broader societal consequences of the pandemic has revealed many difficult trade-offs between health and the economy, and between the needs of different patient groups within the health care sector. While the current attention to such matters is unprecedented, policy makers are confronted with many of these trade-offs also in non-pandemic times. To make informed decisions on policy options, however, requires decision makers to weigh up health and economic consequences, aiming to ensure maximum benefit or minimal harm. Welfare economic tools like cost-benefit-analysis can aid decision makers in this process by providing relevant and clear information to openly address the nature of the trade-offs being made (Chilton, Nielsen, & Wildman, 2020; Donaldson & Mitton, 2020; Hendren & Sprung-Keyser, 2020).

Cost-benefit analyses entail measuring and valuing gains and losses (benefits and costs) in monetary units, thereby allowing a holistic perspective on societal trade-offs and identifying which policy option is socially most preferred. In the context of interventions and policies affecting population health (though not necessarily aimed primarily at health), cost-benefit analysis therefore requires to obtain estimates on the monetary equivalent value of health, from here onwards denoted as v_Q (McIntosh, 2010).

In the narrow health care context, v_Q, depending on the jurisdiction (Rowen, Azzabi Zouraq, Chevrou-Severac, & van Hout, 2017), constitutes an important parameter in health technology assessment. There, value for money considerations are often operationalised using cost-utility analysis, where a new technology's costs are compared to its expected health gain, measured using Quality Adjusted Life Years (QALYs) (Neumann, Sanders, Russell, Siegel, & Ganiats, 2016). Equation (1) formulates a generalisation of the corresponding decision rule, with ΔQ denoting the health gain (in QALYs) and Δc_t the total costs compared to the alternative treatment:

$$(1) \ \frac{\Delta c_t}{\Delta Q} < v_Q$$

This cost-effectiveness ratio (ICER) is acceptable, if it lies below v_Q corresponding to one QALY (Brouwer, van Baal, van Exel, & Versteegh, 2019).[1] While the use and empirical foundation of such threshold values within health care vary across jurisdictions (Cameron, Ubels, & Norström, 2018; Cleemput, Neyt, Thiry, De Laet, & Leys, 2011), estimating the level of v_Q corresponding to one QALY, also for the purpose of cost-benefit analysis, is challenging and has been attempted using various methods (see background section). In this endeavour, Huang, Frijters, Dalziel, and Clarke (2018) were the first to conceptualise and apply the well-being

---

[1] In the broader cost-benefit framework, this QALY equivalent v_Q value can be used for transforming health gains into monetary benefits.

valuation approach for estimating a QALY equivalent v_Q, providing estimates of A\$42,000 (€28,000) to A\$67,000 (€45,000). This method is based on the marginal rate of substitution between income and health. Further exploration of the approach is needed to be able to judge whether the corresponding estimates are indeed helpful for informing v_Q. This paper aims to make the following contributions: Firstly, by applying a similar approach as Huang et al. (2018) and using data from a different context, we generate further insights regarding the validity and reliability of the well-being valuation method for determining v_Q. Secondly, we aim to address some empirical and methodological challenges associated with applying the well-being valuation method in general and for valuing QALYs in particular, which were not fully addressed in previous studies. By using German data, an additional contribution lies in providing information on v_Q for a context in which such estimates are scarce, a result of German health authorities not (explicitly) basing their reimbursement decisions on the framework outlined in Equation (1).[2]

We used data from the German Socio-Economic Panel (2019), or SOEP, from 2002 to 2018. Fixed-effects and instrumental variable regressions were used to address endogeneity concerns regarding the impact of income on life satisfaction. Our baseline estimates indicate population average monetary valuations of a QALY of €22,717 and €58,533, with and without instrumenting for income. However, alternative specifications and robustness checks lead to varying estimates, highlighting the empirical challenges and the consequences of methodological choices on the obtained monetary values, and areas for future research.

---

[2] Instead, the trade-off between Δc_t and ΔQ is discussed and determined in closed-door price negotiations between health authorities and the manufacturer. The methodological uncertainty around estimating v_Q has been cited as a key reason for the scepticism towards adopting more transparent threshold-based decision rules (Bundesministerium für Gesundheit, 2008).

2 The search for v_Q and the well-being valuation method

Various methods have been used in the ongoing endeavour of obtaining estimates of v_Q, producing a range of conceptually different values. One approach, employed by Mason, Jones-Lee, and Donaldson (2009), bases v_Q on estimates of the value of preventing a statistical fatality, a concept commonly used in public sector safety policies. Another approach calculating v_Q entails using relative risk aversion in relation to income (Phelps, 2019). However, v_Q estimates have predominantly been obtained based on stated preferences, by asking individuals directly about their willingness to pay (WTP) for specific health gains. Ryen and Svensson (2015) summarised the extensive literature that used WTP methods to identify v_Q and reported trimmed mean and median estimates of €74,159 and €24,226 (in 2010 price levels).

Huang et al. (2018) proposed an alternative method for estimating v_Q, based on revealed, although subjective, information: the well-being valuation approach. This method has been applied to obtain monetary valuations for various other non-market goods, including specific health outcomes and diseases (Brown, 2015; Ferrer-i Carbonell & van Praag, 2002; Howley, 2017; McNamee & Mendolia, 2018), informal care provision (Mcdonald & Powdthavee, 2018; van den Berg & Ferrer-i Carbonell, 2007), air pollution and natural disasters (Luechinger, 2009; Luechinger & Raschky, 2009), national security (Frey, Luechinger, & Stutzer, 2009) or the welfare effects of sports events (Dolan et al., 2019). In their study, Huang et al. (2018) used data from the HILDA panel survey from Australia and obtained v_Q estimates of A$42,000 (€28,000) to A$67,000 (€45,000), which were similar to threshold values applied for funding decisions in Australia. Recently, Himmler, van Exel, and Brouwer (2020) applied the well-being valuation approach in a cross-sectional sample from the UK to estimate v_Q, as well as an equivalent value for broader well-being. They report a base case v_Q estimate of £30,786 (approximately €35,000).

Both stated preference WTP and well-being valuation approaches have advantages and disadvantages and may answer different questions based on how v_Q is specified. The former allows researchers to tailor their experimental design to specific contexts and control for undesired influences. For instance, WTP can be expressed from an individual or societal perspective (Bobinac, van Exel, Rutten, & Brouwer, 2013), capturing more than self-interested motivations when establishing WTP-based v_Q estimates. Similarly, equity concerns relating to specific health states or streams (Dolan & Olsen, 2001; Pinto-Prades, Sánchez-Martínez, Corbacho, & Baker, 2014), but also socio-economic health inequalities can be connected with the QALY framework (Wagstaff, 1991). Furthermore, one can also pose WTP questions from an *ex-ante* or *ex-post* perspective, with the former having the advantage of capturing options value (Gyrd-Hansen, 2003; Philipson & Jena, 2006).

However, the practice of asking individuals directly for the value of a prospect brings unique challenges; hypothetical response bias and insensitivity to scope or framing effects are only some of the practical concerns (see Kling, Phaneuf, and Zhao (2012)) that have been found to apply when obtaining WTP estimates for a QALY (Ahlert, Breyer, & Schwettmann, 2016; Bobinac, van Exel, Rutten, & Brouwer, 2012; Gyrd-Hansen, Jensen, & Kjaer, 2014; Soeteman, van Exel, & Bobinac, 2017).

The well-being valuation approach avoids these challenges by relying on (usually) large-scale observational data, promising to provide a more inclusive picture of the range of preferences over health and wealth across diverse sub-populations. However, the approach limits the scope to respondents' individual *ex-post* valuations, while endogeneity concerns are a prevailing issue as it relies on the estimation of causal effects of health and income to calculate trade-offs.

3 Methods

3.1 Conceptual framework

We generally followed the framework proposed by Huang et al. (2018) for obtaining v_Q. In a simplified model, the subjective well-being (SWB) of individual i at time T, as a proxy for individual utility, is assumed to be described by:

$$(2)\ W_i = W(Y_{it}, H_{it})$$

Where W_{it} is a vector of the individual's well-being at all observed time points (w_{it}), Y_{it} is the corresponding incomes (y_{it}), and H_{it} a vector of health states (h_{it}). The total well-being experienced by individual i over a time interval of length T can then be described by a simple cumulative sum of individual well-being states across time.

$$(3)\ W_i = \sum_{t=0}^{T} W(Y_{it}, H_{it})$$

Within this framework, consider an individual experiencing a change to their health vector ΔH_i within the time window T. For the individual to remain on the same level of subjective well-being W_i requires an offsetting income change ΔY_i;

$$(4)\ W_i = W(Y_i + \Delta Y_i, H_{it} + \Delta H_i)$$

The proposed approach estimates the population average ΔY necessary to offset an imposed hypothetical health state change ΔH over T equivalent to one QALY. Therefore, ΔY is the compensating income variation for one QALY, or short CIV_{QALY}.

3.2 Baseline specification

Following Huang et al. (2018), an ordinary least squares (OLS) fixed-effects regression was estimated to calculate the impact of health and income on SWB within a time window T of two years (t_0 and t_{-1}). Modelling SWB as linear despite the cardinal nature of life satisfaction is a widely used approach, see e.g. Ferrer-i Carbonell and van Praag (2002). The underlying empirical model takes the following form;

$$(5)\ W_{irt} = \alpha + \beta_0 H_{irt} + \beta_1 H_{irt\text{-}1} + \delta_0 Y_{irt} + \delta_1 Y_{irt\text{-}1} + \tau X_{irt} + \lambda_i + \mu_r + \epsilon_t + u_{irt}$$

where W_{irt} refers to the subjective well-being of individual i living in region r at time t, measured using life satisfaction data. The individual's health status H_{irt} is captured by health utility values based on the short form six dimensions (SF-6D) instrument and its UK utility tariff (Brazier & Roberts, 2004). Household income is denoted by Y_{irt}. Lagged variables of health and income were included to not be limited to short-term one-year changes and to partly account for reverse causality. We control for a vector X_{irt} of other potential time-varying confounders. To account for time-invariant unobservables, we incorporated individual (λ_i), state (μ_r), and time (ϵ_t) fixed effects. u_{irt} denotes the error term. Heteroscedasticity-robust standard errors were used in all estimations.

In a second step, we obtained CIV_{QALY} values by dividing the health status coefficients (β_0 and β_1) by the income coefficients (δ_0 and δ_1):

$$(6)\ CIV_{QALY} = \frac{\beta_0+\beta_1}{\delta_0+\delta_1}$$

The corresponding values represent the marginal rate of substitution between income and health with respect to well-being, based on the overall population average. CIV_{QALY} thereby is the empirical conceptualisation of v_Q using the well-being valuation approach. Income outliers (as will be defined in section 2.4) were dropped from the baseline analysis.

3.3 Instrumental variable specification

A well-documented problem of the well-being valuation approach is the endogeneity of the income coefficient estimate. This was frequently addressed using an instrumental variable (IV) (see e.g. Howley (2017), McNamee and Mendolia (2018), and Brown (2015)). Huang et al. (2018) instrumented income with the occurrence of financial-worsening-events such as personal bankruptcy or large financial losses.

Lacking such information, we followed Luechinger (2009), who used predicted labour-market earnings based on industry-occupation cells as income instrument. The rationale is that shifts in predicted income correspond to industry and/or occupation wide trends, which correlate with the development of negotiated wages or collective wage agreements, but do not reflect individual level effort or circumstances. Further, it is assumed that the income variance across industries and occupations captures information on the unobserved costs of income generation such as stress and/or associated health risks, and that unobserved selection effects of certain types of individuals into industries and occupations are captured in the time-invariant fixed-effects. One advantage of this instrument is that the captured income shifts have a rather permanent nature, whereas financial-worsening-events or lottery wins can be highly transitory shocks. In addition, permanent income shifts have been found to be of higher relevance for individuals' well-being (Bayer & Juessen, 2015; Cai & Park, 2016).

The identifying assumption is, therefore, that income variation across industries and occupations over time is uncorrelated with individual-level characteristics and especially life satisfaction, besides the effect of income changes themselves. To implement the IV approach, we followed a two-stage least squares estimation procedure. In a first step we estimated the individual's labour vmarket earnings L_{irt} based on the following regression;

$$(7)\ L_{irt} = \alpha + p_0 I_{irt} + p_1 O_{irt} + p_2 T_{irt} + p_3 R_{irt} + \mu_r + \epsilon_t + u_{irt}$$

from which we obtained fitted values, constituting the predicted labour earning conditional on the individual's industry-occupation cell (I_{irt} and O_{irt}), work tenure (T_{irt}), and workhours (R_{irt}) and a set of industry- and year-fixed-effects.

The obtained predicted labour earnings were summed on the household level and weighted by household composition to obtain the predicted household labour income $\hat{L}^{HH}_{irt}$, the instrument used in the first-stage regression;

$$(8)\ Y_{irt} = \alpha + \bar{\beta}_0 H_{irt} + \bar{\beta}_1 H_{irt\text{-}1} + \bar{\delta}_0 \hat{L}^{HH}_{irt} + \bar{\delta}_1 \hat{L}^{HH}_{irt\text{-}1} + \bar{\tau} X_{irt} + \bar{\lambda}_i + \bar{\mu}_r + \bar{\epsilon}_t + \bar{u}_{irt}$$

from which we obtained the fitted values for individual income, $\hat{Y}_{irt}$. In the second stage we substituted income Y_{irt} by $\hat{Y}_{irt}$, estimating

$$(9)\ W_{irt} = \alpha^I + \beta^I_0 H_{irt} + \beta^I_1 H_{irt\text{-}1} + \delta^I_0 \hat{Y}_{irt} + \delta^I_1 \hat{Y}_{irt\text{-}1} + \tau^I X_{irt} + \lambda^I_i + \mu^I_r + \epsilon^I_t + u^I_{irt}$$

The resulting coefficients for health (β_0 and β_1) and income (δ_0 and δ_1) were then included in Equation (6) to calculate the IV CIV_{QALY} estimate. For further details please see Online Supplement O2.A3.

3.4 Alternative model specifications

Treatment of outliers

Due to a right-skewed and long-tailed income distribution, with self-reported income often misreported or even exaggerated (Hariri & Lassen, 2017), income outliers may have a large effect on CIV_{QALY} estimates when using linear models (Rousseeuw & Leroy, 1987). To identify outliers, which remains challenging for fixed-effects models (Verardi & Croux, 2009), we reformulated our base case model as a pooled OLS model and calculated DFbeta, a measure quantifying the impact that dropping an observation has on the coefficient estimate. All observations with a DFbeta larger than 1, the recommended threshold (Bollen & Jackman, 1985), were dropped from the baseline analysis. In a robustness check we repeated the calculations including these outliers.

Income specification

To accommodate the diminishing marginal return of income we log-transformed income (Layard, Nickell, & Mayraz, 2008). CIV_{QALY} was then estimated based on a slightly modified equation as used by Ólafsdóttir, Ásgeirsdóttir, and Norton (2020) and van den Berg and Ferrer-i Carbonell (2007). This entailed dropping the lagged income and health coefficients as used in our base model (Equation 6).

$$(10)\ CIV_{QALY} = \bar{y} * \left(\exp\left(\frac{-\beta_0 * \frac{1}{\Delta}}{\delta_0} \right) - 1 \right) * \Delta$$

In the log-income specification CIV_{QALY} was calculated as the percentage share of annual income (median annual income y). By construction, CIV_{QALY} values would be confined to be no greater than this income level which may be acceptable when valuing small gains or changes but not a full QALY. Therefore, we added the parameter Δ to the equation and set it to 10. Instead of calculating the monetary equivalent of a one QALY change we calculated the equivalent of a 0.1 QALY change and multiplied it by 10.

To account for the non-linearity of income without imposing a logarithmic functional form, which may not adequately capture the relationship especially on the lower end of the income distribution, we furthermore tested a piecewise linear specification similar to Ólafsdóttir et al. (2020). To obtain the appropriate number of income splines and cut-off values, we iteratively combined income-deciles. The equality of coefficient estimates of adjacent splines was tested and non-significantly different splines were gradually combined until coefficients were significantly different and model fit did not improve. CIV_{QALY} values were then calculated for each income spline and aggregated by weighting according to the number of individuals in the respective splines. Estimating a piecewise IV specification was not feasible, as one distinct income instrument would have been required for each of the splines.

Choice of utility tariff

Lacking a German specific SF-6D utility tariff we relied on the UK time-trade-off based value set (Brazier & Roberts, 2004) to construct health utilities. In an alternative specification we explored the importance of tariff choice by instead applying a recently developed value set from the Netherlands which was estimated using a discrete choice experiment (Jonker, Donkers, de Bekker-Grob, & Stolk, 2018).

Health state dependence of the utility of consumption

Another empirical issue of concern relates to the interaction between health and income and experienced (consumption) utility. This so-called health state dependence implies that the marginal utility gain from a given income change is directly dependent on the underlying health status (Finkelstein, Luttmer, & Notowidigdo, 2013). So far, there is only inconclusive evidence on the magnitude and the direction of this effect: Finkelstein et al. (2013) found a negative health state dependence, a higher marginal utility of income in good compared to bad health, based on US data. However, replicating their approach using European data, Kools and Knoef (2019) found evidence for positive health state dependence, potentially due to differing provision of public goods in European healthcare systems.

As illustrated by both Finkelstein et al. (2013) and Kools and Knoef (2019), health state dependence has important implications for (health) economic issues such as the optimal design of insurance contracts or individual-level decisions on life-cycle savings. In the context of estimating CIV_{QALY}, which requires a simultaneous measurement of the well-being impacts of both health and income separately, a thorough investigation of the life-cycle development of health states and the associated changes in consumption utility seems warranted.

To explore the potential impact of health state dependence on CIV_{QALY} estimates, we reduced our sample to those individuals that transitioned between health states. Finkelstein et al. (2013) used the onset of chronic diseases for this purpose. While this represents a convenient definition for an elderly population, we took a different approach, allowing us to observe the transition of individuals from good to bad health also for healthier groups. First, we reduced the sample to individuals whose mental or physical short form health questionnaire (SF-12) component scores changed by at least 10, or one standard deviation, throughout their respective observation period.[3] This was done to ensure that individuals in this group have experienced a consequential change in their mental and/or physical health. Good health states were defined as periods in which either of the two scores was above their respective individual-level mean; bad health states if they were below. Secondly, we conditioned on the consecutive observation of differing health states with at least two consecutive periods needed to be observed in either state. This allowed us to estimate CIV_{QALY} for good and bad health separately while also ensuring that individuals transition into longer-term health states (see Online Supplement O2.A4 for details). Importantly, the sample included individuals transitioning from good to bad health and vice versa, although the former is most frequent.

---

3 The SF-12 is also used to calculate SF-6D health utilities. Component scores range from 0 (worst) to 100 (best) with a normalised mean of 50 and standard deviation of 10 (Ware, Keller, & Kosinski, 1995).

4 Data

We used data from the annual SOEP panel survey, providing a representative sample of the adult (aged 16+) German population (Goebel et al., 2019). Ethical approval with respect to the surveying process generating the underlying data was obtained by the SOEP researchers directly. SF-6D health utilities were constructed from SF-12 data, which is biennially included in the survey since 2002. To facilitate the specified two-year time-frame T used for the CIV_{QALY} calculations, and to prevent dropping observations from every second year, we linearly imputed SF-6D values for intermediate years. However, this was only done if individuals were observed for three consecutive years with two completed SF-12 surveys.

Life satisfaction was measured on a 10-point scale ranging from 0 (*"completely dissatisfied"*) to 10 (*"completely satisfied"*). Information on individuals' income was based on self-reported monthly net household income. To account for differences in household composition, we calculated equivalised household income, following the definition by Hagenaars, De Vos, Asghar Zaidi, et al. (1994). Income data was converted to 2018 prices using the official consumer price indices (Federal Statistical Office, 2020).

To construct our instrument, we extracted information on net labour income and individuals' industry and occupation. We dropped households with individuals where information on labour income but not on industry/occupation was available. Predicted labour income was assumed to be zero for all individuals with no labour income information, or who stated that they were not employed.[4]

We furthermore extracted information on a similar set of variables as used by Huang et al. (2018) to control for confounding factors. These included age, disability, marital status, employment status, educational attainment, and leisure time. Table 1 summary statistics of the analysis data, consisting of 29,735 individuals providing 186,906 individual-year observations. Online Supplement Table O2.A1.1 provides an overview of the conditioning applied to the SOEP data, while Table O2.A1.2 shows that the sub-sample of employed individuals who were dropped because of missing industry/occupation information is comparable to the remaining sample of employed individuals. As the exclusion of individuals without at least two consecutive SF-6D values was the only major selection criterion, the sample remained largely representative for the overall German population.

---

[4] Following Luechinger (2009) we added a constant of e1 to all incomes for the log-income specification.

Table 1: Descriptive statistics

Variable	Mean	Std. Dev.	Description
Life satisfaction	7.09	1.71	0 (lowest) to 10 (highest)
Income in 1000's	2.03	1.29	Monthly household income in €
SF-6D utility	0.73	0.13	0.345-1, 1 perfect health
Disability	0.14	0.35	1 if disability status
Age in years	53.67	15.78	
(de facto) Married	0.67	0.47	1 if married, living together
Education: Primary	0.12	0.32	1 if primary educated
Education: Secondary	0.63	0.48	1 if secondary educated
Education: Tertiary	0.25	0.43	1 if tertiary educated
Leisure time	2.18	2.03	Hours per day
Employed	0.56	0.50	1 if employed
Unemployed	0.04	0.21	1 if unemployed
Work hours	21.22	20.99	Hours per week
Tenure	7.03	9.96	Years at current job
Observations		186,902	
Individuals		29,735	

Source: Own calculations based on SOEP Waves 2002-2018.

5 Results

5.1 Baseline results

The baseline OLS and IV results, are shown in Table 2, separating between results using the full dataset with imputed SF-6D values, and the dataset without imputation. To construct our instrumental variables, we predicted labour incomes based on industry/occupation for 125,229 observations. Online Supplement O2.A3 provides details on this prediction and the associated errors, which were small for the largest part of the income distribution. The instruments were significant in the first stage regression (Online Supplement Table O2.A3.1) and passed the Cragg-Donald weak identification test (F-value: 1,864 and 192). This indicates a high relevance of the instrument, a common finding for this type of instrument (Bayer & Juessen, 2015; Luechinger, 2009). The Hausman test for endogeneity of the instrumented variables was significant, signalling that income should not be treated as exogenous.

Equivalised monthly household income, health status (SF-6D utility), and their lagged values were positive and significant predictors of life satisfaction in the OLS specification. This was also the case when instrumenting for income, except that the lagged income coefficient was insignificant. We observed a two-fold increase in the income coefficients in the IV model (0.048 vs. 0.098), a similar magnitude to what has been observed in previous studies using the SOEP (Bayer & Juessen, 2015; Pischke, 2011). Interestingly, the difference is minimal compared to what was observed by Huang et al. (2018), who reported an IV coefficient which was 130 times larger than the OLS coefficient (0.080 and 0.0006). Applying the estimated income and SF6D coefficients to Equation (6) resulted in a CIV_{QALY} value of €58,533 in the OLS model and €22,717 when instrumenting for income. This value represents the average amount of additional income necessary to maintain the same level of life satisfaction if a hypothetical health change of one QALY is imposed.

Without SF-6D imputation, reducing our sample to 85,433 observations across 21,718 individuals, the OLS results increased by a factor of 1.38 to €80,522 while the IV-based value increased by a factor of 1.24 to €28,130. These differences were driven by larger SF-6D and income coefficients compared to the baseline calculations, possibly resulting from increased within-person variance and time-frame T being two years instead of one. For the remainder of the results presented, we will be using the full dataset with imputed SF-6D values to make use of the largest amount of information available.

Table 3 columns 2-3 contains estimates for East and West Germany separately, motivated by the persisting differences in life satisfaction and income levels (Frijters, Haisken-DeNew, & Shields, 2004; Vatter, 2020). OLS-based CIV_{QALY} estimates were €75,748 in the West and €28,548 in the East. The IV-based estimate was also higher in the West compared to the East (€20,750 and €12,982), although the relative difference was lower (factor of 3.64 and 2.20). In both models, this difference was mainly driven by a considerably larger income coefficients in the East, likely due to the prevailing income differences between West and East; observed average monthly equivalised income was €2,140 in the West and only €1,652 in the East.

We investigated the (undesired) impact of macro-economic conditions on CIV$_{QALY}$ estimates by excluding the years of the financial crisis and recession in Germany (2007-2009). As shown in Table 4 (columns 4-6), this had only a minor impact on the OLS and IV CIV$_{QALY}$ values (€54,567 and €20,574). However, estimates based on the pre-crisis time periods 2002-2006 (€56,640 and €7,720) were substantially lower compared to estimates based on data from 20102018 (€70,572 and €24,811). This resulted from larger estimated effects of income in earlier periods, which may both be a result of a positive trend in incomes or a shift in population preferences and values over the last decades. Online Supplement Table O2.A2.1 provides further results on age and gender subgroups.

Table 2: Baseline Results

	SF-6D Imputation				**No Imputation**			
	OLS		IV		OLS		IV	
Income in 1000's	0.05***	(0.01)	0.10***	(0.03)	0.05***	(0.01)	0.14***	(0.05)
Income (t– 1)	0.01	(0.01)	0.04	(0.03)	-0.00	(0.01)	-0.00	(0.07)
SF-6D utility	3.12***	(0.06)	3.12***	(0.05)	3.52***	(0.06)	3.51***	(0.05)
SF-6D utility (t– 1)	0.10*	(0.06)	0.10*	(0.05)	0.47***	(0.05)	0.46***	(0.05)
disability	-0.14***	(0.02)	-0.14***	(0.02)	-0.09***	(0.03)	-0.09***	(0.02)
Age	0.09***	(0.01)	0.08***	(0.02)	0.05***	(0.01)	0.05***	(0.01)
Age squared	-0.00***	(0.00)	-0.00**	(0.00)	-0.00**	(0.00)	-0.00	(0.00)
(de facto) Married	0.18***	(0.02)	0.18***	(0.02)	0.17***	(0.03)	0.16***	(0.02)
Primary education	-0.18*	(0.09)	-0.21***	(0.08)	-0.10	(0.15)	-0.13	(0.13)
Tertiary education	-0.18***	(0.06)	-0.19***	(0.05)	-0.19***	(0.07)	-0.20***	(0.07)
Leisure time	0.03***	(0.01)	0.03***	(0.00)	0.03***	(0.01)	0.03***	(0.01)
Leisure time squared	-0.00***	(0.00)	-0.00***	(0.00)	-0.00**	(0.00)	-0.00***	(0.00)
Unemployed	-0.52***	(0.03)	-0.53***	(0.02)	-0.53***	(0.04)	-0.53***	(0.03)
Work hours	0.00***	(0.00)	0.00***	(0.00)	0.00***	(0.00)	0.00	(0.00)
Tenure	-0.01***	(0.00)	-0.01***	(0.00)	-0.01***	(0.00)	-0.01***	(0.00)
Model statistics								
Cragg-Donald			1,864				192	
Anderson			3,642				382	
Endogeneity test			10.0				5.8	
BIC	540,754		540,995		250,099		236,538	
Observations	186,902		186,902		93,450		85,433	
Individuals	29,735		29,735		29,735		21,718	
CIV$_{QALY}$ in €	58,533		22,717		80,522		28,130	

Source: Own calculations based on SOEP Waves 2002-2018. * p < 0.10, ** p < 0.05, *** p < 0.01. BIC Bayesian information criteria.

Table 3: Results by region and time-period

	Baseline		**East**		**West**		**w/o 2007-2009**		**2002-2006**		**2010-2018**	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Income in 1000's	0.05^{***}	0.10^{***}	0.13^{***}	0.18^{**}	0.04^{***}	0.07^{**}	0.05^{***}	0.11^{***}	0.06^{***}	0.29^{***}	0.04^{***}	0.09^{*}
	(0.01)	(0.03)	(0.02)	(0.08)	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)	(0.09)	(0.01)	(0.05)
Income in 1000's	0.01	0.04	0.00	0.03	0.01	0.04	0.01^{*}	0.05	-0.00	0.10	0.01	0.04
(t– 1)	(0.01)	(0.03)	(0.02)	(0.06)	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.08)	(0.01)	(0.04)
SF-6D utility	3.12^{***}	3.12^{***}	2.90^{***}	2.90^{***}	3.18^{***}	3.17^{***}	3.16^{***}	3.15^{***}	2.93^{***}	2.92^{***}	3.08^{***}	3.08^{***}
	(0.06)	(0.05)	(0.13)	(0.12)	(0.07)	(0.07)	(0.07)	(0.07)	(0.15)	(0.15)	(0.08)	(0.08)
SF-6D utility (t– 1)	0.10^{*}	0.10^{*}	-0.12	-0.12	0.16^{**}	0.16^{**}	0.10	0.09	0.06	0.06	-0.07	-0.07
	(0.06)	(0.05)	(0.12)	(0.12)	(0.07)	(0.07)	(0.07)	(0.07)	(0.14)	(0.14)	(0.08)	(0.08)
Model statistics												
Cragg-Donald		1,863.7		323.9		680.2		783.4		181.2		494.3
Anderson		3,642.0		544.4		1,265.5		1,429.5		328.8		907.3
Endogeneity test		10.0		1.5		5.8		9.7		8.2		2.7
BIC	540,755	540,995	127,072	127,092	412,723	412,877	431,238	431,487	129,869	130,432	276,374	276,464
Observations	186,902	186,902	43,447	43,447	143,361	143,361	151,461	151,461	48,678	48,678	101,048	101,048
CIV_{QALY} in €	58,533	22,717	20,750	12,982	75,748	28,548	54,567	20,574	56,640	7,720	70,572	24,811

Source: Own calculations based on SOEP Waves 2002-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. BIC Bayesian information criteria.

5.2 Specifications related to income

Re-estimating our baseline models including four individual-year observations which were flagged as outliers lead to a considerably lower income coefficient in the OLS model (Table 4 columns 3-4). This increased the CIV_{QALY} value to €82,484. The IV estimates were only minimally affected by this (€22,782). The outlier observations corresponded to two individuals from the same household, which reported a drop in monthly income from €142,534 to €14,051 within two consecutive years, while reporting constant life satisfaction.

In the models using log-transformed income (Table 4 columns 5-6), the income coefficient was 0.24, larger than reported before by Pischke (2011) (0.125 to 0.182), who also used the SOEP. The corresponding IV coefficient, with a value of 0.63, was on the higher end of previous IV estimates based on the industry-wage structure and the SOEP: Luechinger (2009) reported an estimate of 0.55, while Pischke (2011) reported values ranging from 0.489 to 0.617. Previous estimates based on instruments using lagged or future income shocks were also similar, with Bayer and Juessen (2015) providing a range of 0.45 to 0.50 for permanent income shifts.[5] The log-transformation resulted in considerably larger CIV_{QALY} values compared to the baseline. The OLS values increased by a factor of 2.63 to €153,877 while the IV values increase by a factor of 3.59 to €81,649.[6]

The piecewise linear specification was estimated with ultimately four income splines. The cut-off points were at the 20*th* percentile (€1,200), the 40*th* percentile (€1,546), and the 80*th* percentile (€2,635). Figure 1 plots the overall distribution of life satisfaction across income, and the linear fit of life satisfaction across splines, indicating a non-linear, diminishing pattern. The spline specific CIV_{QALY} values were €7,347, €11,686, €29,548, and €409,810. The population aggregated CIV_{QALY} was €97,486. This estimate was driven by the large CIV_{QALY} value in the fourth income spline, where the income coefficient was insignificant. Using the three significant splines lead to a CIV_{QALY} value of €19,515.

---

[5] Bayer and Juessen (2015) used only data from West Germany, possibly leading to a downward bias due to higher income levels in the West. Similarly, both Pischke (2011) and Luechinger (2009) use SOEP waves from the years before the East German SOEP sample was established in 1990 alongside waves containing samples from both former German states past 1990.

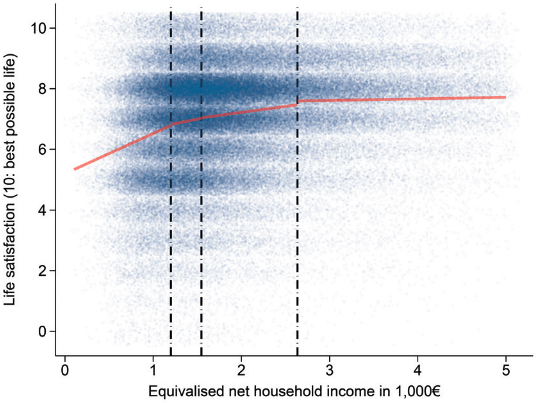
[6] Huang et al. (2018) did not observe a large difference between linear and log income-based estimates. However, they multiplied the ratio of income and health coefficients as in Equation (6) with the median income to obtain CIV_{QALY} (as opposed to Equation (10)).

Table 4: Income Specifications

	Baseline		**Without outliers**		**Log income**		**Piece-wise**
	OLS	IV	OLS	IV	OLS	IV	OLS
Income in 1000's	0.05^{***}	0.10^{***}	0.03^{***}	0.10^{***}			
	(0.01)	(0.03)	(0.01)	(0.03)			
Income in 1000's (t– 1)	0.01	0.04	0.01^{***}	0.04			
	(0.01)	(0.03)	(0.00)	(0.03)			
SF-6D utility	3.12^{***}	3.12^{***}	3.12^{***}	3.12^{***}	3.18^{***}	3.16^{***}	3.18^{***}
	(0.06)	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
SF-6D utility (t– 1)	0.10^{*}	0.10^{*}	0.10^{*}	0.10^{*}			
	(0.06)	(0.05)	(0.06)	(0.06)			
Log income					0.24^{***}	0.63^{***}	
					(0.02)	(0.13)	
1st income							0.43^{***}
spline							(0.05)
2nd income							0.27^{***}
spline							(0.05)
3rd income							0.11^{***}
spline							(0.02)
4th income spline							0.01
							(0.01)
Model statistics							
Cragg-Donald		1,863.7		825.8		1,329.9	
Anderson		3,642.0		1,529.4		1,278.2	
Endogeneity test		10.0		12.9		9.7	
BIC	540,755	540,995	540,801	541,306	540,506	541,501	540,448
Observations	186,902	186,902	186,906	186,906	186,902	186,902	186,902
CIV_{QALY} in €	58,533	22,717	82,484	22,782	153,877	81,649	97,486
w/o 4th spline	540,755	540,995	127,072	127,092	412,723	412,877	19,515

Source: Own calculations based on SOEP Waves 2002-2018. *Note:* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. BIC Bayesian information criteria. Instrumental variable did not pass weak identification tests for piecewise income specification. CIVs for piecewise regression represents population-weighted averages of all splines or the first three splines (€7,347, €11,686, €29,548 and €409,810).

Figure 1: Relationship between life satisfaction and income across income splines



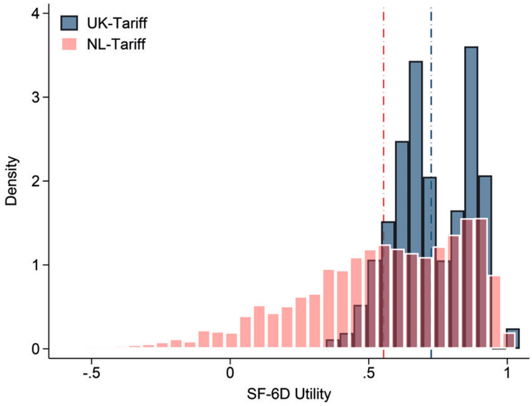
Source: Own calculations based on SOEP Waves 2002-2018. *Note:* Life satisfaction values are depicted as small blue dots. Black dash-dotted vertical lines represent the income splines used in the piece-wise linear regression. Red horizontal lines plot the linear fit within these splines.

5.3 Specifications and issues related to health

Choice of SF-6D value set

Applying the Dutch SF-6D value set shifted the distribution of health utilities (Figure 2), with the mean utility decreasing from 0.725 to 0.554. These differences likely reflect methodological differences rather than actual variation in health state preferences between the UK and the Netherlands as UK and Dutch tariffs for the EQ-5D have been shown to be similar (Norman et al., 2009).

Figure 2: SF12 index values using UK and Dutch tariffs



Source: Own calculations based on SOEP Waves 2002-2018. *Note:* The blue dash-dotted line indicates the Dutch tariff mean. The red dash-dotted line indicates the UK tariff mean. The distributions and means reflect SF-6D values based on self-reported SF12 questionnaires only.

The estimated CIV_{QALY} values using the Dutch SF-6D tariff were markedly smaller (Table 5). The OLS estimates decreased from €58,533 to €32,534, while the IV estimates decreased from €22,717 to €13,054. This shift was caused by the smaller SF-6D coefficients (3.12 to 1.78), resulting from the wider spread of the Dutch tariff, which ranges from -0.44 to 1, allowing for negative health state utility, instead of 0.345 to 1 as in the UK value set. The same actual change in health corresponds to a larger change in SF-6D utility in the Dutch tariff which reduces the impact of a (hypothetical) one unit change in SF-6D on life satisfaction.

Table 5: Choice of SF-6D tariff

	UK tariff		**Dutch tariff**	
	OLS	IV	OLS	IV
Income in 1000's	0.05^{***}	0.10^{***}	0.05^{***}	0.09^{***}
	(0.01)	(0.03)	(0.01)	(0.03)
Income in 1000's (t– 1)	0.01	0.04	0.01	0.05^{*}
	(0.01)	(0.03)	(0.01)	(0.03)
SF-6D utility	3.12^{***}	3.12^{***}	1.78^{***}	1.78^{***}
	(0.06)	(0.05)	(0.03)	(0.03)
SF-6D utility (t– 1)	0.10^{*}	0.10^{*}	0.05	0.05
	(0.06)	(0.05)	(0.03)	(0.03)
Model statistics				
Cragg-Donald		1,863.7		907.1
Anderson		3,642.0		1,671.4
Endogeneity test		10.0		9.4
BIC	540,755	540,995	538,297	538,297
Observations	186,902	186,902	186,906	186,906
CIV_{QALY} in €	58,533	22,717	32,534	13,054

Source: Own calculations based on SOEP Waves 2002-2018. *Note:* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. BIC Bayesian information criteria.

Health state dependence of the utility of consumption

We explored the potential impact of health state dependence on CIV_{QALY} estimates by restricting our sample to individuals experiencing a substantial health change and splitting their respective observation periods into good and bad health states (see section 2.4). The resulting sample was considerably smaller, including only 5,112 individuals yielding 48,861 observations. Nevertheless, the summary statistics suggests that the sample is still comparable to the full population sample (see Online Supplement Table O2.A4.1). Table 6 depicts the corresponding estimation results. Compared to the baseline estimates using the full sample, CIV_{QALY} values based on the combined good and bad health state samples were lower in the OLS model (€39,482) and similar in the IV specification (€20,377). For "good health states", the corresponding CIV_{QALY} estimates were lower with €33,336 and €16,532. For "bad health states", the OLS-based CIV_{QALY} estimate was €38,374 and the IV-based estimate €11,779.

Table 6: Health State Dependence

	Baseline		**Good Health**		**Bad health**	
	OLS	IV	OLS	IV	OLS	IV
Income in 1000's	0.07^{***}	0.17^{**}	0.05^{***}	0.11	0.08^{**}	0.32
	(0.01)	(0.07)	(0.02)	(0.08)	(0.04)	(0.24)
Income in 1000's	0.03^{**}	0.02	0.03^{**}	0.05	0.03	0.05
(t– 1)	(0.01)	(0.06)	(0.01)	(0.06)	(0.03)	(0.17)
SF-6D utility	3.62^{***}	3.60^{***}	2.51^{***}	2.50^{***}	4.10^{***}	4.03^{***}
	(0.11)	(0.09)	(0.14)	(0.12)	(0.38)	(0.37)
SF-6D utility (t– 1)	0.10	0.11	0.12	0.12	0.32	0.32
	(0.10)	(0.10)	(0.12)	(0.11)	(0.26)	(0.27)
Model statistics						
Cragg-Donald		620.7		425.1		95.9
Anderson		1,208.4		828.1		188.4
Endogeneity test		3.0		1.8		1.0
BIC	150,481	150,558	102,463	102,497	37,832	37,899
Observations	48,861	48,861	35,401	35,401	13,460	13,460
CIV_{QALY} in €	39,482	20,377	33,336	16,532	38,374	11,779

Source: Own calculations based on SOEP Waves 2002-2018. *Note:* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. BIC Bayesian information criteria

Important to note is that the drop in the IV based results for the bad health state primarily resulted from a larger income coefficient estimate, even though the SF-6D coefficients increased considerably. These results indicate that there is a positive health state dependence of income in line with the results for Germany by Kools and Knoef (2019). Unfortunately, we were not able to follow Kools and Knoef (2019) and Finkelstein et al. (2013) in focusing on non-working individuals to ensure stable income across health states, ruling out that the increased income coefficients are driven by individuals losing their income, and hence having a larger marginal utility of additional earnings. For our analysis, such a restriction was not feasible, as within person income variation is necessary to estimate the income coefficients. However, the general empirical pattern remains the same when excluding individuals with large negative income differences between health states (see Online Supplement Table O2.A4.2). This also holds when only considering the working population (Table O2.A4.3) and those experiencing sudden and severe health changes (Table O2.A4.4).

5.4 Robustness checks

Lastly, we tested the robustness of our baseline results to some general concerns regarding our estimation strategy (Table 7). In a first robustness check, we limited our sample to individuals which were in paid employment and provided industry-occupation information, the same sample which was used to obtain estimates for predicted labour income for the IV regression. The resulting OLS-based CIV_{QALY} was slightly lower than the baseline at €52,829, while the

IV-based value was slightly higher than the baseline at €26,097. These differences were driven by the smaller SF-6D coefficients in both OLS and IV models, likely resulting from the the working population being healthier as individuals without labour income (the unemployed and retired). The sum of both income coefficients was smaller in the corresponding IV-calculations compared to baseline, increasing the CIV_{QALY}.

Next, we followed Luechinger (2009) by excluding households with self-employed main income earners, as the income measurement error was likely to be amplified among these individuals. Self-employed individuals are often reluctant to disclose their income, while also experiencing unstable income streams and hence, even if not reluctant to report, they might simply misreport accidentally. The resulting CIV_{QALY} estimates and income and SF-6D coefficients were similar to the baseline estimates (€55,359 and €20,352).

Another concern relating to the instrument is that observed income changes may also relate to individual effort, which likely impacts income differently across industries and occupations. Unfortunately, effort cannot be observed. To nevertheless explore this, we use information on reported bonuses, gratifications, or profit sharing to identify the group of individuals for whom this might be a relevant concern, as for them effort would have the highest impact on income and life satisfaction. To test the robustness of our results to this potential bias, we estimate our baseline models excluding such observations. The results in Table 7 columns 7-8 suggest that this bias is relatively limited.

To investigate the potential impact of dropping employed individuals without industry/ occupation information (as required for constructing the IV), we included those observations in a further robustness check (last column 7). The corresponding OLS estimates for income coefficients and CIV_{QALY} (€62,266) are comparable to our baseline estimates. However, by construction, we cannot confirm this for the IV estimates.

Table 7: Robustness Checks

	Baseline		**Working only**		**No self-employed**		**No bonus income**		**Ind/occ**
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS
Income in 1000's	0.05***	0.10***	0.05***	0.05	0.07***	0.05	0.05***	0.14***	0.04***
	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)
Income in 1000's (t– 1)	0.01	0.04	0.01	0.07**	0.00	0.08**	0.01	0.02	0.01**
	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)
SF-6D utility	3.12***	3.12***	2.95***	2.94***	2.97***	2.97***	3.12***	3.11***	3.14***
	(0.06)	(0.05)	(0.08)	(0.07)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)
SF-6D utility (t– 1)	0.10*	0.10*	0.07	0.06	0.01	0.01	0.10*	0.11*	0.12**
	(0.06)	(0.05)	(0.07)	(0.07)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)
Model statistics									
Cragg-Donald		1,863.7		1,355.7		1,898.4		719.1	
Anderson		3,642.0		2,637.7		3,633.0		1,334.4	
Endogeneity test		10.0		5.4		7.4		10.1	
BIC	540,755	540,995	319,169	319,323	279,896	280,043	502,827	503,172	578,002
Observations	186,902	186,902	116,125	116,125	101,703	101,703	172,998	172,998	198,950
CIV$_{QALY}$ in €	58,533	22,717	52,829	26,097	44,058	21,382	53,974	20,464	62,266

Source: Own calculations based on SOEP Waves 2002-2018. *Note:* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. BIC Bayesian information criteria. Ind/occ refers to specification where individuals without industry/occupation information were included.

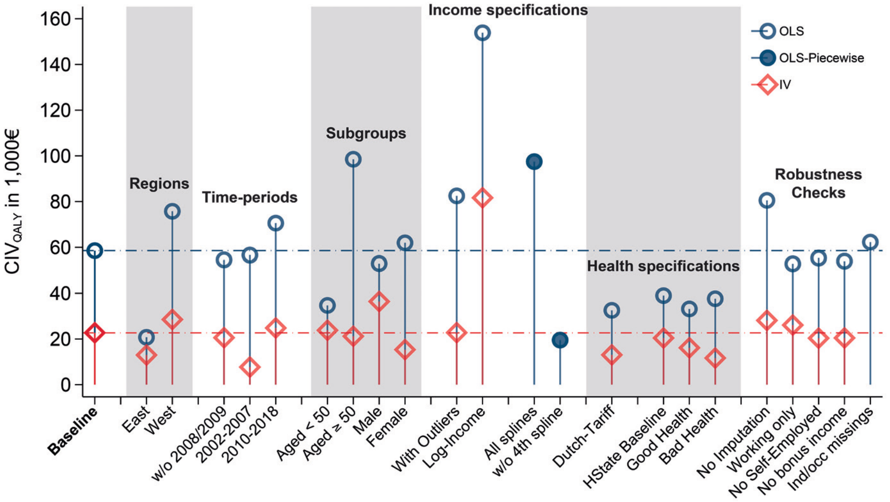
6 Discussion

Applying the well-being valuation approach to longitudinal health and income data from Germany, we estimated the monetary equivalent value of one year in full health v_Q (equivalent to one QALY). Beyond demonstrating the feasibility of this approach in a new country context, we explored additional empirical and methodological challenges with implications for the practical usefulness of well-being valuation based v_Q estimates (denoted as CIV$_{QALY}$).

6.1 Overview and context of results

Figure 3 presents an overview of our CIV$_{QALY}$ estimates. The baseline calculations provided average monetary valuations of a QALY of €58,533 (OLS) and €22,717 (IV). CIV$_{QALY}$ estimates varied across model specifications with the bulk of values lying between €20,000 and €60,000 and the (OLS) log-income specifications reaching the maximum value of €153,877. Instrumenting for income consistently lead to lower values, a common finding in the well-being valuation literature (e.g., Ólafsdóttir et al. (2020)).

Figure 3: Overview of CIV$_{QALY}$ estimates



Source: Own calculations based on SOEP Waves 2002-2018. *Note:* The horizontal dash-dotted lines indicate our baseline CIV$_{QALY}$ estimates from the baseline OLS (blue) and IV (red) specifications.

The range of CIV$_{QALY}$ estimates obtained in our study fit into the ballpark of more reasonable stated preference estimates (Ryen & Svensson, 2015). Furthermore, it is important to note that all IV CIV$_{QALY}$ estimates, except the log-income specification, fell within the range of v_Q estimates for Germany of €4,988 to €43,115 reported by Ahlert et al. (2016), who provided the only v_Q estimates estimates until now. A first approximation of an *opportunity cost based QALY threshold value*, or k_Q, for Germany was reported by Woods, Revill, Sculpher, and Claxton (2016). Using empirical estimates of health care opportunity costs for Germany, and the relationship between

GDP per capita and the value of a statistical life, they calculated a k_Q range of €19,276 to €24,374 (in 2018 euros). A recent related study by Ochalek and Lomas (2020) reported estimates of cost per DALY averted (essentiallly the reciprocal of a QALY gain) for Germany of €47,116 to €74,650 (in 2018 euros).

6.2 Limitations and strengths of the analysis

IV-based estimates rely on restrictive assumptions relating to their unbiasedness and informational value. A valid concern is that occupational choice may be related to other unobserved confounders, such as personality traits or income preferences (Pischke & Schwandt, 2012). The use of individual fixed effects should somewhat alleviate such concerns due to the rather stable nature of personality traits (Borghans, Duckworth, Heckman, & ter Weel, 2008), but they cannot provide complete assurance. A further assumption is that being employed in a certain industry/occupation should not have a significant, direct effect on life satisfaction, therefore violating the exclusion restriction. Online Supplement Tables O2.A3.6 and O2.A3.7 show that, controlling for income and other confounders, this effect is not zero, but modest and mostly insignificant. One additional drawback that is rarely explicitly discussed but of great importance in the well-being valuation context, is that IV estimates only yield a local average treatment effect (Angrist, Imbens, & Rubin, 1996). Using predicted labour income as an instrument, at least questions the generalisability of our IV estimates to the full, also non-working, population. Further, as we are not able to address all sources of measurement error with respect to income, the remaining upward bias in the income coefficients would imply a downward bias in the estimated CIV_{QALY} values.

In addition, income variation in industry-occupation cells predominantly consists of *positive, upward shifts in wages* (and differences therein). This is conceptually different to financial worsening events, as used by Huang et al. (2018), as these capture *income losses*.[7] Given income loss aversion (Boyce, Wood, Banks, Clark, & Brown, 2013), our IV based CIV_{QALY} estimates likely represent a lower-bound.

The potential endogeneity of health (status) in life satisfaction regressions due to reverse causality (see e.g., Veenhoven (2008) or Sabatini (2014)), which is rarely addressed in the related literature, is a further limitation. This endogeneity could be addressed by appropriate instruments or identifying health shocks which are plausibly exogenous, such as heart attacks or strokes. However, besides practical issues like data availability, it is questionable how generalisable such localised causal effects would be for the overall impact of the multi-dimensional construct of health on life satisfaction. Heterogeneity may exist both concerning the type of health shocks, but also relating to their timing within the (life cycle) health distribution. Whether or not our estimates of the impact of health are biased upwards or downwards can therefore not be easily ascertained. In the one previous article in the related literature that addressed endogeneity

---

[7] Ambrosio, Clark, and Zhu (2018) report a persistent direct effect of financial worsening (and improvement) events on life satisfaction beyond income-changes, raising concerns on the general appropriateness such events as income instruments.

directly, Brown (2015) found that the health coefficient was slightly overestimated when not instrumented. Assuming this also holds in our context, this would imply that there is an upward bias in our CIV_{QALY} values resulting from the endogeneity of health.

A more practical limitation relating to measuring health was that we had to impute SF-6D utilities for every second year to make full use of the SOEP's rich annual data. This required us to condition the sample on individuals who had at least three consecutive observations, which may have resulted in underestimating the impact of deteriorating health, since individuals are more likely to discontinue their participation in a longitudinal survey following a negative health shock.

A final limitation lies in the potential presence of double-counting as subjective well-being enters the model twice: As an implicit consideration in the SF-6D health state valuation tasks (on which the scoring of our health measure is based on), and as a proxy for experienced utility (Equation 2). To what extent this is problematic is difficult to assess. To avoid this double counting, one could use an unweighted sum score of the SF-6D levels. However, this raises the question of the appropriate anchoring. Using such a sum score, rescaled to a 0 to 1 range (expanding the number of levels of the first two SF-6D dimensions to five to not impose any weighting) lead to lower CIV_{QALY} estimates in the unimputed dataset (Online Supplement Table O2.A2.2, columns 4-5). However, when imposing the same anchor and therefore range as in the original SF-6D tariff (0.345 to 1), the OLS and IV results (€88,867 and €30,567) were much closer to the unimputed baseline estimates (€80,671 and €27,777).

It seems that not the differential weighting between the dimensions caused the larger differences, but the different anchors, i.e., the lowest utility. Another alternative approach entailed eliciting CIV values for different dimensions directly by regressing on all levels of the SF-6D, which did not impose any weighting. Adding up the resulting CIV values of the lowest level of all six dimensions, summed up to a cumulative value of moving from the best possible to the worst possible health state of €79,013 and €27,489, which again resembled the unimputed baseline estimate (Table A2.2). While these sensitivity checks somewhat alleviate the concerns about double-counting, the latter revealed that 46% of the CIV_{QALY} value stemmed from the impact of mental health on life satisfaction. It is likely that the mental health dimension also plays a dominant role in our baseline calculations. Whether this is problematic lies outside the scope of this paper, as it relates to a more general issue of the well-being valuation approach: is life satisfaction the best (available) proxy for experienced utility?

6.3 Implications of findings

There are several practical implications of our study for future applications of the well-being valuation approach in general, and its use for estimating v_Q in particular. First, judging from the impact outliers have in the OLS specification (Table 4), subsequent applications of the approach using linear models should report on the occurrence and treatment of outliers. Secondly, given that the functional form of income had a large impact on our estimates its final specification must be well argued and reporting results for other alternative functional forms seems warranted.

The piecewise linear specification seems to be a promising alternative, given that it is more flexible and gives all income groups a proportional weight. This approach, however, comes at the price of increasing the number of variables that need to be instrumented for.

Third, the choice of utility tariffs for the health instrument matters greatly. Especially the range of the scoring algorithm has a large impact (Online Supplement Table O2.A2.2), as an imposed one unit change in health utility implies a different change in health if the range goes from 0.345 to 1 or -0.44 to 1. How to overcome this issue while facilitating cross-country comparisons and how this relates to the underlying QALY concept, should further be discussed in future applications. Lacking country specific tariffs, it may be convenient to opt for a tariff whose origin can be placed in cultural and socio-economic proximity to the country to be investigated. However, the impact of methodological peculiarities in how these tariffs were generated are relevant. It would have been interesting also to compute CIV_{QALY} estimates based on the more widely used EQ-5D health utilities and compare the implications of differences in scope and range of the health instrument used on CIV_{QALY} values. Unfortunately, EQ-5D is rarely included longitudinal surveys. Lastly, the differing values obtained when considering East and West Germany separately, or specific time periods (Table 3), also highlight the potential importance of the specific country context for CIV_{QALY} calculations.

One of the major conceptual issues discussed in our analysis, with direct relevance for the practical value of any empirically estimated CIV of health, is the health state dependence of utility. We attempted to provide indicative evidence on how health state dependence might affect estimated CIV_{QALY} values. However, it remains unclear whether empirical approaches based on self-reported (panel) data can produce reliable estimates if health state dependence is prevalent and survey participation and attrition is (partially) driven by health changes over time. We found considerable differences in the estimated CIV_{QALY} values when comparing periods of good and bad health within individuals (Table 6). As the underlying point estimates depicted substantial uncertainty, these findings should be interpreted with caution and merely as indicative evidence for the role of health state dependency in this context. The impact of this sub-sample of individuals on the population wide CIV_{QALY} value is likely small, as attrition is high once individuals experience bad health states, long-term or very severe health shocks. Hence, a pragmatist might argue that this issue is of theoretical interest only. We would argue, however, that this is an inherent limitation of self-reported observational data and its *ex-post* perspective in this context. Stated preference methods would allow for an explicit *ex-ante* consideration of this issue through tailored sampling strategies and survey design.

An additional conceptual concern related to health state dependence is the question of adaptation to bad health over time (Huang et al., 2018). Adaptation implies the gradual return of subjective well-being to pre-health-shock levels despite continued (or deteriorating) bad health (Loewenstein & Ubel, 2008). This phenomenon has been documented before using the SOEP data (Oswald & Powdthavee, 2008) and would generally decrease estimated CIV_{QALY}, as the marginal utility of health would decrease with time spent in bad health. To what extend this represents an estimation error, however, is debatable and depends on what is perceived to be the *"true"* impact of ill-health on well-being over time, and whether adaptation, if present, should be corrected for. The recent findings by Etilé, Frijters, Johnston, and Shields (2020), who documented a heterogeneous distribution of adaptive potential across subgroups, underline the relevance of this concern also from a normative perspective.

The previous remarks highlight avenues for future research, like investigating the causal effect of health on life satisfaction, for example using instrumental variable regressions. In addition, the approach would crucially benefit from further research into the impact of income on life satisfaction, for example using (natural) experiments. The regular inclusion of variables that represent valid instruments for income into different population panel surveys could also be beneficial for further exploring the reliability and validity of these instruments and the approach, as it would allow cross-national replications of results. Meanwhile, future applications may draw upon recent advances into the generalisability of IV-based estimates (see e.g., Mogstad, Santos, and Torgovitsky (2018)) to explore how these concerns can be addressed within the framework of available instruments. Further, linking survey data on individual-level subjective well-being measures with detailed administrative records on income, health, and care consumption would also be a fruitful direction for further inquiry, resolving some of the enumerated concerns. With respect to the question of health state dependency, for example, it would be possible to determine the extent to which survey data has an inherent blind spot due to the attrition of individual following severe health shocks. In addition, such data could also be used to explore a wider range of specification choices within the general empirical strategy used, for example with respect to the choice of control variables. Here, we deliberately followed Huang et al. (2018), as the set of basic control variables they propose is available in most national panel surveys, which facilitates replications across country-contexts. However, there is ample room for extending the analysis by considering a wider set of control variables and their impact on CIV_{QALY} estimates, or even to altogether choose a different approach such as shrinkage estimators (e.g., LASSO) or matching to address endogeneity concerns around the impact of health and/or income on life satisfaction.

A final issue concerns the practical application of our v_Q estimates. If certain (health) policies/interventions in Germany were to be evaluated using a v_Q value from our study, which range from around €20,000 (IV) to €60,000 (OLS), we have to highlight the following:[8] Our study cannot provide a definite answer regarding which estimate is most accurate to be used in different contexts. This relates to the uncertainty surrounding these estimates and the underlying assumptions, but also to normative or distributional questions, which need to be addressed in the future (Cookson et al., 2020). While our piecewise regression results somewhat reflect such concerns by constructing v_Q estimates using a weighted mean of the different parts of the income distribution, this is only a first, very simplistic approach. When used in a normative context, like decisions on reimbursement of technologies, explicit policy (debate and) support is required.

---

[8] Health care funding decisions in Germany are not based on cost utility analysis, partially because thresholds were difficult to define (Bundesministerium für Gesundheit, 2008). Finding comparable monetary estimates using a compensating income variation and stated preference studies to some extent puts this into perspective.

Applied studies could use the range we provided to highlight the impact of varying v_Q estimates on their results and recommendations, keeping in mind that for specific sub-populations our v_Q estimates might not be directly applicable. In any case the selection of any specific value over another in any practical application should be transparently discussed with respect to the applied selection criteria.

7 Conclusions

We demonstrated that the well-being valuation approach *can* be another useful instrument in the (health) economist's toolbox for obtaining monetary equivalent valuations of health (v_Q). Some inherent empirical and conceptual challenges of applying this approach in this context can be addressed, especially when using large-scale longitudinal data. However, other issues, like the health state dependence of the utility of consumption, will remain a threat to the validity of estimates, warranting additional research. Concurrently, alternative approaches of estimating v_Q, like stated preference studies or methods aiming at eliciting the value of a statistical life, as recently applied by Herrera-Araujo, Hammitt, and Rheinberger (2020), provide important complementary insights, despite their conceptual differences. Also given their respective strengths and limitations, methodological diversity is desired in the ongoing endeavour of measuring the monetary equivalent value of health.

The type of v_Q estimates provided in our analysis reflect average marginal health valuations (with the caveat of being entirely based on marginal changes in health-related quality of life), representative on a national level. As such, these can be applied in economic evaluations informing decision making on a societal level for publicly funded policies or interventions. Such v_Q estimates predominantly find their use by informing the cost-effectiveness threshold in the context of cost-utility analysis within health care, which aid in informing decisions on reimbursement of certain health interventions. However, estimates of the monetary value of health can also be useful in broader contexts, like cost-benefit analyses or similar approaches (Cookson et al., 2020), especially when benefits and costs of policies/interventions constitute a mix of health and non-health outcomes occurring across different sectors. Advancing methodologies aiming to estimate v_Q and providing insights into their validity can assist in informing some of the uncomfortable trade-offs that societies generally face in priority-setting both within health care but also beyond (Chilton et al., 2020).

References

Ahlert, M., Breyer, F., & Schwettmann, L. (2016, feb). How you ask is what you get: Framing effects in willingness-to-pay for a QALY. *Social Science & Medicine*, 150, 40–48.

Ambrosio, C. D., Clark, A., & Zhu, R. (2018). Living in the Shadow of the Past : Financial Profiles , Health and Well-Being Living in the Shadow of the Past : Financial Profiles , Health and Well-Being . *Working paper*.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434), 444–455.

Bayer, C., & Juessen, F. (2015). Happiness and the persistence of income shocks. *American Economic Journal: Macroeconomics*, *7*(4), 160–187.

Bobinac, A., van Exel, N. J. A., Rutten, F. F., & Brouwer, W. B. (2012). Get more, pay more? an elaborate test of construct validity of willingness to pay per qaly estimates obtained through contingent valuation. *Journal of Health Economics*, 31(1), 158–168.

Bobinac, A., van Exel, N. J. A., Rutten, F. F. H., & Brouwer, W. B. F. (2013, oct). Valuing QALY gains by applying a societal perspective. *Health Economics*, 22(10), 1272–1281.

Bollen, K. A., & Jackman, R. W. (1985, may). Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. *Sociological Methods & Research*, 13(4), 510–542.

Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43(4), 972–1059.

Boyce, C. J., Wood, A. M., Banks, J., Clark, A. E., & Brown, G. D. A. (2013). Money, well-being, and loss aversion: Does an income loss have a greater effect on well-being than an equivalent income gain? *Psychological Science*, 24(12), 2557-2562.

Brazier, J. E., & Roberts, J. (2004). The Estimation of a Preference-Based Measure of Health From the SF-12. *Medical Care*, *42*(9), 851–9.

Brouwer, W., van Baal, P., van Exel, J., & Versteegh, M. (2019). When is it too expensive? Cost-effectiveness thresholds and health care decision-making. *European Journal of Health Economics*, 20(2), 175–180.

Brown, T. T. (2015). The Subjective Well-Being Method of Valuation: An Application to General Health Status. *Health Services Research*, 50(6), 1996–2018.

Bundesministerium für Gesundheit. (2008). *Stellungnahme zur Methodik der Kosten-Nutzen-Bewertung von Arzneimitteln*.

Cai, S., & Park, A. (2016). Permanent income and subjective well-being. *Journal of Economic Behavior & Organization*, 130, 298–319.

Cameron, D., Ubels, J., & Norstr¨om, F. (2018). On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. *Global Health Action*, 11(1).

Chilton, S., Nielsen, J. S., & Wildman, J. (2020). Beyond COVID-19: How the ‘dismal science’can prepare us for the future. *Health Economics*, 29(8), 851.

Cleemput, I., Neyt, M., Thiry, N., De Laet, C., & Leys, M. (2011, jan). Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *International Journal of Technology Assessment in Health Care*, 27(1), 71–76.

Cookson, R., Skarda, I., Cotton-Barratt, O., Adler, M., Asaria, M., & Ord, T. (2021). Quality adjusted life years based on health and consumption: A summary wellbeing measure for cross-sectoral economic evaluation. *Health Economics*, 30(1), 70-85.

Dolan, P., Kavetsos, G., Krekel, C., Mavridis, D., Metcalfe, R., Senik, C., ... Ziebarth, N. R. (2019). Quantifying the intangible impact of the olympics using subjective well-being data. *Journal of Public Economics*, 177, 104043.

Dolan, P., & Olsen, J. A. (2001). Equity in health: the importance of different health streams. *Journal of Health Economics*, 20(5), 823–834.

Donaldson, C., & Mitton, C. (2020). Coronavirus: Where has all the health economics gone? *International Journal of Health Policy and Management*, 9(11), 466–468.

Etil´e, F., Frijters, P., Johnston, D. W., & Shields, M. (2020). Psychological resilience to major socioeconomic life events. *IZA Discussion Paper*, *No. 13063*.

Federal Statistical Office. (2020). *Consumer price index (incl. rates of change): Germany, years. Table: 61111-0001.*

Ferrer-i Carbonell, A., & van Praag, B. M. (2002, dec). The subjective costs of health losses due to chronic diseases. An alternative model for monetary appraisal. *Health Economics*, 11(8), 709–722.

Finkelstein, A., Luttmer, E. F., & Notowidigdo, M. J. (2013). What good is wealth without health? the effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, 11(suppl 1), 221–258.

Frey, B. S., Luechinger, S., & Stutzer, A. (2009). The life satisfaction approach to valuing public goods: The case of terrorism. *Public Choice*, 138(3-4), 317–345.

Frijters, P., Haisken-DeNew, J. P., & Shields, M. A. (2004). Money does matter! evidence from increasing real income and life satisfaction in east germany following reunification. *American Economic Review*, 94(3), 730–740.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schr¨oder, C., & Schupp, J. (2019). The german socio-economic panel (soep). *Jahrbücher für Nationalökonomie und Statistik*, 239(2), 345–360.

Gyrd-Hansen, D. (2003). Willingness to pay for a qaly. *Health Economics*, 12(12), 1049–1060.

Gyrd-Hansen, D., Jensen, M. L., & Kjaer, T. (2014). Framing the willingness-to-pay question: impact on response patterns and mean willingness to pay. *Health Economics*, 23(5), 550– 563.

Hagenaars, A. J., De Vos, K., Asghar Zaidi, M., et al. (1994). *Poverty statistics in the late 1980s: Research based on micro-data*. Office for Official Publications of the European Communities.

Hariri, J. G., & Lassen, D. D. (2017). Income and Outcomes: Social Desirability Bias Distorts Measurements of the Relationship between Income and Political Behavior. *Public Opinion Quarterly*, 81(2), 564–576.

Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3), 1209–1318.

Herrera-Araujo, D., Hammitt, J. K., & Rheinberger, C. M. (2020). Theoretical bounds on the value of improved health. *Journal of Health Economics*, 102341.

Himmler, S., van Exel, J., & Brouwer, W. (2020). Estimating the monetary value of health and capability well-being applying the well-being valuation approach. *European Journal of Health Economics*, 21, 1235–1244.

Howley, P. (2017). Less money or better health? Evaluating individual's willingness to make trade-offs using life satisfaction data. *Journal of Economic Behavior & Organization*, 135, 53–65.

Huang, L., Frijters, P., Dalziel, K., & Clarke, P. (2018). Life satisfaction, QALYs, and the monetary value of health. *Social Science & Medicine*, 211(June), 131–136.

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018). Advocating a Paradigm Shift in Health-State Valuations: The Estimation of Time-Preference Corrected QALY Tariffs. *Value in Health*, 21(8), 993–1001.

Kling, C. L., Phaneuf, D. J., & Zhao, J. (2012). From Exxon to BP: Has some number become better than no number. *Journal of Economic Perspectives*, 26(4), 3–26.

Kools, L., & Knoef, M. (2019). Health and consumption preferences; estimating the health state dependence of utility using equivalence scales. *European Economic Review*, 113, 46–62.

Layard, R., Nickell, S., & Mayraz, G. (2008). The marginal utility of income. *Journal of Public Economics*, 92(8-9), 1846–1857.

Loewenstein, G., & Ubel, P. A. (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*, 92(8-9), 1795–1810.

Luechinger, S. (2009). Valuing Air Quality Using the Life Satisfaction Approach. *The Economic Journal*, 119, 482–515.

Luechinger, S., & Raschky, P. A. (2009). Valuing flood disasters using the life satisfaction approach. *Journal of Public Economics*, 93(3-4), 620–633.

Mason, H., Jones-Lee, M., & Donaldson, C. (2009, aug). Modelling the monetary value of a QALY: a new approach based on UK data. *Health Economics*, 18(8), 933–950.

Mcdonald, R., & Powdthavee, N. (2018). The Shadow Prices of Voluntary Caregiving: Using Panel Data of Well-Being to Estimate the Cost of Informal Care. *IZA Discussion Paper Series* (11545).

McIntosh, E. (2010). Introduction. In E. McIntosh, P. Clarke, E. Frew, & J. Louviere (Eds.), *Applied methods of cost–benefit analysis in health care* (pp. 3–18). Oxford University Press.

McNamee, P., & Mendolia, S. (2019). Changes in health-related quality of life: a compensating income variation approach. *Applied Economics*, 51(6), 639-650.

Mogstad, M., Santos, A., & Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, 86(5), 1589–1619.

Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (2016). *Cost effectiveness in health and medicine*. New York: Oxford University Press.

Norman, R., Cronin, P., Viney, R., King, M., Street, D., & Ratcliffe, J. (2009). International Comparisons in Valuing EQ-5D Health States :. *Value in Health*, 12(8), 1194–1200.

Ochalek, J., & Lomas, J. (2020). Reflecting the Health Opportunity Costs of Funding Decisions Within Value Frameworks: Initial Estimates and the Need for Further Research. *Clinical Therapeutics*, 42(1), 44–59.e2.

Olafsd´ottir, T.,´ Asgeirsd´ottir, T. L., & Norton, E. C. (2020). Valuing pain using the subjective´ well-being method. *Economics and Human Biology*, 37.

Oswald, A. J., & Powdthavee, N. (2008). Does happiness adapt? a longitudinal study of disability with implications for economists and judges. *Journal of Public Economics*, 92(5-6), 1061–1077.

Phelps, C. E. (2019). A New Method to Determine the Optimal Willingness to Pay in CostEffectiveness Analysis. *Value in Health*, 22(7), 785–791.

Philipson, T. J., & Jena, A. B. (2006). Who benefits from new medical technologies? estimates of consumer and producer surpluses for hiv/aids drugs. In *Forum for Health Economics & Policy* (Vol. 9).

Pinto-Prades, J.-L., S´anchez-Mart´ınez, F.-I., Corbacho, B., & Baker, R. (2014). Valuing qalys at the end of life. *Social Science & Medicine*, 113, 5–14.

Pischke, J.-S. (2011). Money and hapiness: Evidence from the industry wage structure. *NBER Working Paper* (17056).

Pischke, J. S., & Schwandt, H. (2012). A cautionary note on using industry affiliation to predict income. National Bureau of Economic Research. *NBER Working Paper* (w18384).

Rousseeuw, P. J., & Leroy, A. M. (1987). Robust regression and outlier detection (Vol. 1). Wiley Online Library.

Rowen, D., Azzabi Zouraq, I., Chevrou-Severac, H., & van Hout, B. (2017). International Regulations and Recommendations for Utility Data for Health Technology Assessment. *PharmacoEconomics*, 35(s1), 11–19.

Ryen, L., & Svensson, M. (2015). The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Economics*, 24(10), 1289–1301.

Sabatini, F. (2014). The relationship between happiness and health: evidence from italy. *Social Science & Medicine*, 114, 178–187.

Socio-Economic Panel. (2019). *Socio-economic panel (soep), data for years 1984–2018, version 35.* SOEP.

2

Soeteman, L., van Exel, J., & Bobinac, A. (2017). The impact of the design of payment scales on the willingness to pay for health gains. *The European Journal of Health Economics*, 18(6), 743–760.

Van den Berg, B., & Ferrer-i-Carbonell, A. (2007). Monetary valuation of informal care: the well-being valuation method. *Health Economics*, 16(11), 1227-1244.

Vatter, J. (2020). Well-being in germany: What explains the regional variation? *SOEPpapers*, *No. 435*.

Veenhoven, R. (2008, sep). Healthy happiness: effects of happiness on physical health and the consequences for preventive health care. *Journal of Happiness Studies*, 9(3), 449–469.

Verardi, V., & Croux, C. (2009, sep). Robust Regression in Stata. *The Stata Journal: Promoting communications on statistics and Stata*, 9(3), 439–453.

Wagstaff, A. (1991). QALYs and the equity-efficiency trade-off. *Journal of Health Economics*, 10(1), 21–41.

Ware, J. E., Keller, S. D., & Kosinski, M. (1995). *Sf-12*: How to score the sf-12 physical and mental health summary scales. Health Institute, New England Medical Center.

Woods, B., Revill, P., Sculpher, M., & Claxton, K. (2016). Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. *Value in Health*, 19(8), 929–935.

3

CHAPTER 3

Adaptation in Life Satisfaction and Self-Assessed Health to Disability – Evidence from the UK

Based on: Stöckel, J., van Exel, J., and Brouwer, W. B. (2023). Adaptation in Life Satisfaction and Self-Assessed Health to Disability – Evidence from the UK. *Social Science & Medicine,* 328, 115996.

Abstract:

Experiencing deteriorating health has implications for your quality of life. The theory of adaptation suggests that with time spend living in a health state individuals can adapt, resulting in observed quality of life levels to revert or stagnate despite persistently decreased health. Adaptation has implications for the use of subjective quality of life indicators when quantifying the impact of health changes or the benefits from new medical technologies. As both the impact from ill health and the benefit from new interventions might be disease- or subgroup-specific adaptation further raises ethical concerns but empirical evidence on its existence, magnitude, and heterogeneity remains inconclusive. This paper uses a general population sample of 9,543 individuals that participate in the UK Understanding Society survey and experience the onset of a long-standing illness or disability to provide evidence on these questions. Using ordered-response fixed effects models we explore longitudinal changes in self-assessed health and life satisfaction around the onset of disability. Our results indicate that disability onset is associated with large decreases in subjective health and well-being. Over time this initial decrease in subjective quality of life indicators attenuates, especially in life satisfaction and to a lesser extent for self-assessed health. While the relative difference in adaptation across these two measures remains persistent, we find that across demographic and severity groups the initial impact of disability onset and adaptation differs considerably in its magnitude. These results have important implications for studies aiming to quantify the impact of health conditions on quality-of-life outcomes, especially when using observational datasets.

1 Introduction

Adaptation to ill health refers to the phenomenon that individuals over time adjust to a deterioration in their health (Frederick & Loewenstein, 1999). Adaptation may lead to an increase in reported levels of quality of life, ceteris paribus, even if the health status has not improved. This may be observed when measuring quality of life using self-assessed health or multi-item health state evaluations (Groot, 2000), but also when using broader subjective well-being measures, which are increasingly seen as a relevant maximand for heath and public policy (Benjamin et al., 2019; Frijters, Clark, Krekel, & Layard, 2020; Loewenstein & Ubel, 2008; Peasgood, Foster, & Dolan, 2019). This has consequences for the use of subjective quality of life measures in health economic evaluations and health policy. For instance, the fact that economic evaluations often use members of the general public rather than patients as the source for health state valuations was importantly inspired by evidence of adaptation (Brazier et al., 2018; Frederick & Loewenstein, 1999; Versteegh & Brouwer, 2016) and the need to protect patients from the negative consequences of adaptation in allocation decisions (Cohen, 1993; Menzel et al., 2002). After all, if adaptation leads to higher reported subjective health or well-being, the potential gains from health and social care interventions become smaller.

While there is theoretical motivation for the importance of adaptation in the context of subjective quality of life measures as maximands in health economics and health policy the empirical findings regarding its existence and magnitude are not always conclusive. Since Brickman et al. (1978) first documented patterns consistent with adaptation among paraplegics, several studies have explored the dynamic impact of ill health on self-reported well-being and health outcomes. Most of these studies have explored adaptation using longitudinal panel surveys, allowing for the observation of individuals transitioning from good into ill health, and the application of panel fixed-effects approaches. Lucas (2007) and Powdthavee (2009) found conflicting evidence for adaptation to the onset of self-reported disability using the same German panel data. More recently Ta (2019) considered the differential impact of mental versus physical disability on life satisfaction conditional on an initial loss in life-satisfaction with evidence suggesting adaptation to be specific to the nature of health changes and the considered subgroup. Binder and Coad (2013), McNamee and Mendolia (2014), Cubí-Mollá et al. (2017) and Baji and Bíró (2018) on the other hand focused on the onset of specific conditions e.g., hypertension or diabetes, and report various levels of adaptation depending on the data, approach and definition of ill health considered. Most recently, de Hond et al. (2019) and Bussière, Sirven, and Tessier (2021) explored adaptation using the same pan-European panel survey of older individuals but come to somewhat diverging conclusions based on different empirical strategies. Both report adaptation with respect to life satisfaction but the results by Bussière et al. (2021) suggest these findings to be sensitive to the choice of subjective wellbeing measures considered.

This study explores adaptation to ill health using a sample of 9,235 individuals transitioning into living with a long-standing illness or disability participating in waves 1-10 of the UK longitudinal survey Understanding Society (University of Essex Institute for Social and Economic Research, 2020). Following de Hond et al. (2019), we consider adaptation in life satisfaction and self-assessed health. Life satisfaction is an interesting outcome due to its characteristics as an easy-to-measure conceptualisation of subjective well-being with high policy relevance (Frijters et al., 2020). It encompasses various quality of life dimensions (de Hond et al., 2019) and is highly predictive of individuals' decision making (Kaiser & Oswald, 2022), making it an attractive and relevant experienced utility measure (Dolan & Metcalfe, 2012). It has also received considerable attention in health policy and economics (see e.g., Dolan and Kahneman (2008) or Peasgood et al. (2019)). A growing literature for example has used well-being valuation methods to provide estimates for the monetary equivalent value of health changes to inform health economic decision making, ranging from specific conditions (Howley, 2017; Ólafsdóttir, Ásgeirsdóttir, & Norton, 2020) to generic health changes equivalent with the loss of one quality-adjusted life year (Himmler, Stöckel, van Exel, & Brouwer, 2021; Huang, Frijters, Dalziel, & Clarke, 2018). Especially in this context adaptation in subjective well-being is a relevant concern as such methods rely on the trading-off between the relative impact of health and income changes on life-satisfaction to estimate monetary equivalents.

Exploring adaptation on self-assessed health on the other hand is motivated twofold. Firstly, past studies suggest that adaptation in broader well-being measures is driven by adaptation in subjective health perceptions and the related satisfaction-domain (Powdthavee, 2009). Cross-sectional evidence suggests that also with respect to specific health state evaluations adaptation exists (Groot, 2000; Jonker et al., 2017). Evidence from studies using panel data on the other hand provide a mixed picture with some finding adaptation to occur to a limited degree (Baji & Bíró, 2018; de Hond et al., 2019) while others find no adaptation in self-perceived health or only to occur over decades (Cubí-Mollá et al., 2017). Secondly, while less relevant for this study there is a strong link between economic or health-related decisions and health perceptions, ranging from risky and preventive health behaviours (see e.g. Arni, Dragone, Goette, and Ziebarth (2021); Spitzer and Shaikh (2022)) to decisions on retirement and saving (see e.g., Gan, Gong, Hurd, and McFadden (2015); Schünemann, Strulik, and Trimborn (2017); Spaenjers and Spira (2015)). Adaptation in self-perceived health could indicate a misalignment between objectively experience health changes and the subjective perception of these with potential implications for individual decision making on health.

This study contributes to the literature on adaptation in three ways. Firstly, we study adaptation using a different definition of transitioning into ill health, namely, the onset of a long-standing illness or disability, changes to the short-form 12-item health questionnaire (SF12) mental and physical component scores (Ware, Keller, & Kosinski, 1995), and different levels of functional limitations. This allows us to explore the extent to which adaptation depends on the definition of health and whether it is limited to mild conditions or also observed after more substantial health changes. Secondly, next to de Hond et al. (2019) and Cubí-Mollá et al. (2017) we are only the third study to explore adaptation using (fixed effects) ordered response models. As most studies on adaptation primarily considered life satisfaction or happiness as the main outcome of interest, common practice has been to follow Ferrer-i Carbonell and Frijters (2004) by assuming life satisfaction to be a cardinal measure allowing for regular linear (fixed effects) models to be estimated. An attractive property of such approaches is the ability to directly interpret linear coefficients but recently the appropriateness of such methods to analyse ordinal quality of life data has been debated (see e.g., Kaiser, Vendrik, et al. (2019) or Bond and Lang (2019)). However, ordered response models can be applied using fixed effects designs to obtain policy relevant insights not only into the mean changes on an assumed linear scale but along the distribution of ordered outcomes (Chen, Oparina, Powdthavee, & Srisuma, 2022). Lastly, we explore this topic using the Understanding Society dataset, an annually conducted general population survey. Much of the recent work on adaptation has focused on using surveys among specific populations, such as older individuals (Baji & Bíró, 2018; Bussière et al., 2021; de Hond et al., 2019) or younger cohort studies (Cubí-Mollá et al., 2017). Using a general population survey, we can explore adaptation among a broader range of subgroups.

2 Data

2.1 Data Selection and Definitions

We use data from the United Kingdom Household Longitudinal Study, also known as Understanding Society (USoc). USoc is an annual panel survey covering a representative sample of the adult UK population and is the direct successor to the British Household Panel Survey. Data collection started in 2009 with approximately 50,000 respondents across 30,000 households. USoc applies an overlapping panel design in which each wave is collected over a 24-month period while individuals are interviewed each year. We use the first ten waves of USoc which surveyed individuals between 2009 and 2020 with surveying for wave 10 being completed in the first quarter of 2020.

Our main outcomes of interest are two annually collected subjective quality of life measures, subjective well-being, and health. Subjective well-being is measured using a life satisfaction question asking respondents "how satisfied are you with your life overall?". Individuals respond by indicating on a 7-point scale ranging from 1 (*"completely dissatisfied"*) to 7 (*"completely satisfied"*). Self-assessed health is measured by asking individuals to complete the sentence "in general would you say your health is …" with one of five categories ranging from "poor" to "excellent". To identify individuals transitioning from good into ill health we use the question "do you have any long-standing physical or mental impairment, illness, or disability?". The survey text defines "long-standing" as "anything that has troubled you over a period of at least 12 months or is likely to trouble you for a period of at least 12 months". We use this question to identify those individuals that are observed in good health, not reporting such a long-standing illness or disability (LSI), and a state of ill health, where an LSI is reported for multiple years. We therefore select individuals for the analysis sample based on the observed response pattern across their individual participation waves.

We condition the sample on those respondents providing at least two consecutively observed waves as this is the basic condition for both states to be observable. Secondly, if an individual reports an LSI in the first observed wave we exclude this participant from the analysis as we cannot observe the transition between health states. Likewise, we exclude individuals who never report an LSI, or report an erratic pattern such as multiple spells of periods with and without an LSI. If an individual drops out temporarily due to non-response, we further require this individual to provide at least two consecutive observations on each side of the non-response gap and only allow for one such gap to occur over the available participation waves. In the case where an individual drops out for one wave but then enters the panel again, now reporting a long-standing illness, we assume that the illness onset coincided with the missing wave to calculate the long-standing illness duration. In case an individual reports multiple spells of long-standing illness or disability with a single wave with no LSI in-between, we treat these spells as a period of repeated reporting of ill health indicating a longer-term health change that we consider as a singular spell. For all subsequent results, we have explored whether this choice alters impacts our results, but this decision was found to have no discernible impact. In case an individual stops consistently reporting an LSI, the observations following the LSI spell are removed.

Apart from conditioning on a sample of individuals observed to have transitioned into a state of ill health, we also condition on the availability of control variables at each observed period. These variables were household income, employment status, educational attainment, marital status, and the presence of children in the household. After imposing all of the aforementioned conditions we are left with a sample of 9,543 unique individuals providing 52,553 individual-year observations. The average respondent is observed for six periods equally split between good (no LSI reported) and ill health (LSI reported). Individuals are therefore observed in periods before an LSI is reported ($t < 0$) and after transitioning into worse health ($t \geq 0$). While we are in principle able to observe individuals living with an LSI for up to nine years, the group living with an LSI for six or more years has a small number of individuals observed and are therefore grouped together ($t \geq 5$). Table 1 provides an overview of the number of individuals observed at the different relative time-points to LSI onset for the full sample and by gender and age.

Table 1: Duration Table

	Time to Onset in Years										
	≤-5	-4	-3	-2	-1	0	1	2	3	4	≥5
Total	5,813	3,168	4,277	5,869	9,235	9,235	4,750	3,402	2,381	1,708	2,715
Sex											
Male	2,514	1,374	1,874	2,599	4,040	4,040	2,090	1,497	1,046	749	1,188
Female	3,299	1,794	2,403	3,270	5,195	5,195	2,660	1,905	1,335	959	1,527
Onset Age											
<55	3,199	1,833	2,476	3,440	5,393	5,393	2,532	1,726	1,171	817	1,217
≥55	2,614	1,335	1,801	2,429	3,842	3,842	2,218	1,676	1,210	891	1,498
Observations											52,553
Individuals											9,543

Source: Own calculations based on USoc Waves 2009-2020.

Individuals responding to the LSI question in the affirmative are further asked to indicate whether the reported LSI results in *"substantial difficulties"* in any or multiple of 11 *"areas of life"*. Online Supplement Table O3.A1.2 provides a breakdown of our analysis sample with respect to the dimensions affected and the number of limitations reported. While most respondents (54.49%) do not report the LSI onset to be associated with any dimension to be substantially limited, those respondents that do report such a limitation do so largely in areas of physical mobility and functioning. In our baseline analyses we do not consider this information but when turning towards the role of severity of health changes we do consider this information alongside other indicators to quantify health changes within our sample.

2.2 Summary Statistics

Table 2 provides the summary statistics for the analysis sample alongside the sample of excluded individuals never reporting to have any LSI to illustrate the differences between these groups already before LSI-onset. Note that for the LSI group reported means and standard deviations correspond to the waves prior to onset. The analysis sample was considerably older than the never-LSI sample, with 48 years of age compared to 40. Given this age difference we also observe the analysis sample to be already less healthy before the onset of any LSI across measures of health and well-being. We observe slightly lower life satisfaction and considerably lower self-assessed health levels among respondents in the analysis sample. This is also reflected in the multi-dimensional health measures from the SF12. The SF12 component scores, ranging from 0 (worst) to 100 (best) with a mean of 50 and standard deviation of 10 (Ware et al., 1995), are slightly lower in the mental health dimension and 2 points lower (20% of the standard deviation) for physical health. This is also reflected in the almost twice as large proportion of individuals reporting their health to limit daily activities. With respect to other characteristics, differences are in line with the observed age-difference. The analysis sample contains more retired individuals and has a lower average equivalised income (this measure follows Hagenaars, De Vos, Asghar Zaidi, et al., (1994) by calculating a weighted income per capita measure for each household with the first adult receiving a weight of 1, subsequent adults and children above 13 a weight of 0.5 and other children a weight of 0.3), while also educational attainment is lower, reflecting birth-cohort differences.

Table 2: Summary Statistics

	Analysis Sample (pre LSI onset)		Never LSI Sample	
	Mean	Std. Dv.	Mean	Std. Dv.
Outcomes				
Life Satisfaction	5.29	1.4	5.4	1.35
Self-Assessed Health	3.68	0.88	3.95	0.82
Health Status and Age				
SF12: Mental Component Score	50.12	9.24	50.75	8.5
SF12: Physical Component Score	52.88	7.55	55.02	5.79
Health Limits Typical Activities	0.15	0.36	0.08	0.28
Age in years	48.04	17.23	39.90	15.8
Socio-Economic Controls				
Employed	0.54	0.50	0.61	0.49
Self-Employed	0.08	0.27	0.09	0.28
Unemployed	0.04	0.2	0.04	0.20
Retired	0.22	0.41	0.09	0.29
Working full-time	0.46	0.50	0.54	0.50
Equivalised household income	1713.42	1407.76	1875.22	2995.41
Living in urban area	0.74	0.44	0.76	0.43
Highest Education: Primary/other/none)	0.10	0.30	0.06	0.23
Highest Education: Secondary	0.52	0.50	0.52	0.50
Highest Education: Tertiary	0.38	0.49	0.43	0.49
Single	0.19	0.39	0.29	0.45
Married/Partnership	0.68	0.47	0.64	0.48
Widowed	0.05	0.22	0.02	0.13
Separated/Divorced	0.08	0.27	0.05	0.23
Children living in Household	0.34	0.47	0.43	0.50
Observations	28,362		117,608	
Individuals	9,543		23,395	

Source: Own calculations based on USoc Waves 2009-2020.

3 Methods

In the empirical literature on the determinants of subjective well-being it is common practice to use life satisfaction as a subjective well-being measure assume cardinality despite its categorical nature (Ferrer-i Carbonell & Frijters, 2004). This is attractive as it allows for the application of fixed-effects ordinary least squares models instead of non-linear models for categorical data. This comes with two benefits. Firstly, regression coefficients and effect sizes are easier to interpret due to the linear scale of the dependent variable. Secondly, time-invariant (un)observed factors are absorbed within the individual fixed-effects, already accounting for a large share of confounding variation and allowing for a causal interpretation of estimated coefficients if only this source of variation is of concern. The fixed linear scale also allows for a comparison of effect sizes across models. In non-linear ordered response models, the underlying scale varies across models prohibiting the direct comparisons based on regression coefficients alone.

Past studies of adaptation to ill health and disability have relied on the cardinality assumption, e.g., Oswald and Powdthavee (2008) or Powdthavee (2009). Most of these studies focused on well-being constructs measured using an implied continuous scale such as *"complete dissatisfaction"* to *"complete satisfaction"*. Our empirical strategy is in line with previous work by Cubí-Mollá et al. (2017) and de Hond et al. (2019) who explore the existence of adaptation in self-assessed health and subjective well-being alongside each other by using an ordered logit (fixed-effects) model. Modelling individuals' life satisfaction and self-assessed health using a non-linear ordered response model has multiple benefits. As illustrated by Dickerson, Hole, and Munford (2014), Baetschmann et al. (2015) and Muris (2017) the reliance on linear models for the analysis of categorical data can lead to inconsistent and biased effect estimates. Furthermore, there has been an ongoing debate around how to analyse ordered response quality of life data. The discussion has mostly focused on the question of how such data can be analysed and under what conditions the estimated relationships offer relevant and valid insights (see e.g., Bond and Lang (2019), Kaiser et al. (2019) and Chen et al. (2022) for discussions). In this context ordered response models relying on a dichotomisation to allow for a fixed-effects panel-data based approach, such as the estimator proposed by Baetschmann et al. (2015), have been recommended to provide a practical alternative to linear models (Chen et al., 2022). In linear models' coefficients indicate point changes but for categorical variables these point changes are often uninformative as the underlying linear scale is only assumed while individuals respond by selecting distinct, discrete levels. Non-linear ordered response models take this into account and allow for estimating the change in response probabilities across the observed life satisfaction and self-assessed health distributions.

To do so we apply the *"blow-up-and-cluster"* (BUC) estimator developed by Baetschmann et al. (2015) and implemented in Baetschmann, Ballantyne, Staub, and Winkelmann (2020). Following Baetschmann et al. (2015), assume an individual's subjective well-being and self-assessed health and its determinants can be expressed by the following equation:

$$y^*_{it} = \sum_{j=0}^{5} \gamma_j \, LSI^j_{it} + \sum_{l=1}^{L} \beta_l x^l_{it} + \alpha_i + \epsilon_{it} \text{ with } i = 1,\dots,N \text{ and } t = 1,\dots,T$$

where y^*_{it} is the latent life satisfaction or self-assessed health of a given individual i at time point t. The dummy variables LSI^j_{it} capture the time an individual has spent living with an LSI, ranging from j=0 periods (onset) to j≥0 years of consecutive LSI reporting. Lastly x^l_{it} represents the L time-varying control variables, α_i, the individual fixed-effect and ϵ_{it} the remaining error term.

The observed self-assessed health and life satisfaction values y_{it} are related to the laten variable y^*_{it} based on an observation rule;

$$y_{it} = k \text{ if } \tau_{ik} < y^*_{it} \le \tau_{ik+1} \text{ with } k = 1,\dots,K$$

with individual response thresholds τ_i increasing in K ($\tau_{ik} < y^*_{it} \le \tau_{ik+1} \forall k$) and $\tau_{i0} = -\infty$ and $\tau_{ik+1} = \infty$. Further, the individual error terms ϵ_{it} are assumed to be independent and identically distributed with a logistic cumulative distribution function depending on the individual fixed-effect α_i and a vector of control variables X_{it} which contains the L control variables and the J dummies for LSI duration:

$$F(\epsilon_{it} | X_{it}, \alpha_i) = F(\epsilon_{it}) = \frac{1}{1 + \exp(-\epsilon_{it})} \equiv \Lambda(\epsilon_{it})$$

The probability of observing a specific outcome k for individual i at time t is then given by

$$\Pr(y_{it} = k | X_{it}, \alpha_i) = \Lambda(\tau_{ik+1} - \beta X_{it} - \alpha_i) - \Lambda(\tau_{ik} - \beta X_{it} - \alpha_i)$$

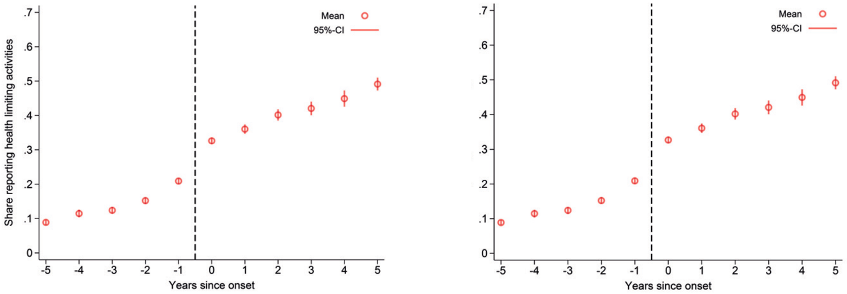
And thereby depends on the vector of estimated coefficients β of the individual control variables X_{it}, the individual fixed-effect α_i, and the individual-specific thresholds τ_{ik} and τ_{ik+1}. However, because for a given observed response only $\tau_{ik} - \alpha_i = \alpha_{ik}$ is identified this results in an identification problem. In addition, in empirical contexts with large N and small T the so-called incidental parameter problem (Lancaster, 2000) leads to inconsistently estimated α_{ik}. The estimator proposed by Baetschmann et al. (2015) solves this issue by collapsing the observed outcomes y_{it} into a set of K binary variables d^k_{it} with $d^k_{it} = 1$ if $y_{it} \geq k$ and then using conditional maximum likelihood estimations for binary outcomes and clustering standard errors on the individual level. This procedure lends the estimator its name (Baetschmann et al., 2015).

4 Results

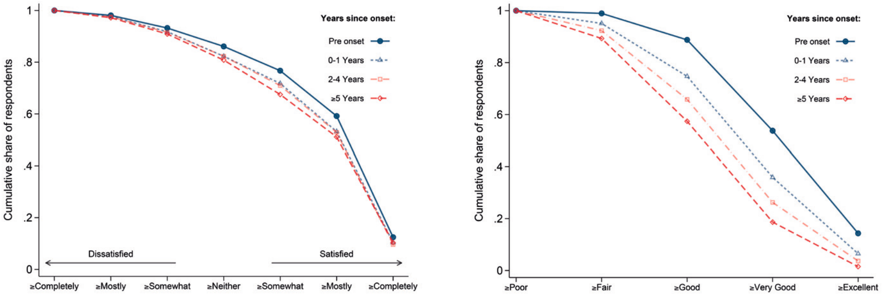
4.1 Descriptive Results

Before considering the results of the ordered response fixed-effects models we provide some descriptive evidence on adaptation. Figure 1 presents information on the health state of the respondents in our analysis sample and their subjective outcome measures around LSI-onset.

Figure 1: Subjective and Objective Health and Well-being Outcomes around the LSI Onset



(a) Mental/Physical Component Score (b) Health Limits Typical Activities



(c) Life Satisfaction (d) Self-Assessed Health

Source: Own calculations based on USoc Waves 2009-2020. *Note:* Panel (a) depicts the average mental and physical component scores by time to LSI-onset, panel (b) depicts the share of individuals reporting to be limited by their health to conduct typical activities. Bars indicate 95-% confidence intervals. The solid vertical line indicates the time point between the last wave directly before a first-reported LSI and the first wave with a reported LSI. Panel (c) depicts the distribution of life satisfaction categories by relative time since LSI-onset and panel (d) depicts the same for self-assessed health. The pre-onset group contains responses from the two years preceding before LSI-onset.

Figure 1 illustrates a couple of key properties of our analysis sample. The onset of an LSI is associated with a decrease in physical health (see Figure 1 panels (a) and (b)). This decrease is already manifested in the period directly preceding the LSI-onset, on the left-hand side of the solid line, while physical health deteriorates further over time. Overall, the onset of an LSI is associated with the largest year-to-year deterioration. For both life satisfaction (panel (c)) and self-assessed health (panel (d)), LSI-onset is associate with an increase in the share of individuals reporting lower levels of these measures. However, for life-satisfaction, there is little difference between the distribution of responses in the first two years of living with an LSI and those living with an LSI for longer periods. For self-assessed health, the distributions indicate that with longer LSI duration individuals more often report lower health. Nonetheless, the cumulative distributions also indicate that the initial negative association is largest in the first two years and that additional years of living with an LSI reporting are associated with smaller decreases despite a continuous deterioration in physical health.

4.2 Baseline Analysis

We further explore the patterns shown in Figure 1 using the fixed-effects ordered logit approach by Baetschmann et al. (2015). Table 3 presents our baseline results, for these we include the coefficients for our control variables in the results table, but for all subsequent results tables in the Online Supplement we instead follow the recommendation to focus only on the parameters of interest (Hünermund & Louw, 2020), in our context the time-dummies for LSI-duration. Doing so cautions against the direct interpretation of the conditional coefficients of control variables included in our analysis to avoid the so-called Table 2 Fallacy (Westreich & Greenland, 2013).

Compared to the reference category, living without an LSI, LSI-onset is associated with a decrease in both life-satisfaction and self-assessed health, as can be seen from the negative coefficients. With increasing time spend living with an LSI the strength of the negative association decreases. For life-satisfaction, the size of coefficients is roughly halved and insignificant for those living with an LSI for three years and more. For self-assessed health, the coefficient size decreases more slowly while the coefficients remain significant. Both patterns would be consistent with individuals adapting over time to living with an LSI and trending towards pre-onset response levels. Although for self-assessed health this adaptation is only partial as the estimated coefficients remain negative and highly significant.

Table 3: Baseline Results

	Life Satisfaction		Self-Assessed Health	
LSI Duration				
Onset	-0.213***	(0.031)	-0.937***	(0.031)
Year 1	-0.230***	(0.046)	-1.031***	(0.046)
Year 2	-0.181***	(0.052)	-0.884***	(0.051)
Year 3	-0.103	(0.065)	-0.798***	(0.061)
Year 4	-0.109	(0.075)	-0.660***	(0.074)
Year 5	-0.112	(0.083)	-0.508***	(0.081)
Control Variables				
Self-employed (*Ref: Employed*)	0.052	(0.078)	-0.005	(0.068)
Unemployed	-0.242***	(0.067)	0.104	(0.066)
Retired	0.414***	(0.080)	0.052	(0.074)
Working full-time	0.179***	(0.048)	0.223***	(0.047)
log(Equivalised Income)	0.062**	(0.021)	0.028	(0.022)
Living in urban area	-0.099	(0.112)	0.179	(0.104)
Education: Secondary (*Ref: Primary/other*)	-0.242	(0.228)	0.318	(0.222)
Education: Tertiary	-0.445	(0.249)	0.319	(0.248)
Married/Partnership (*Ref: Unmarried*)	0.151	(0.084)	-0.014	(0.095)
Widowed	-0.182	(0.169)	-0.035	(0.178)
Separated/Divorced	-0.183	(0.115)	-0.019	(0.124)
Number of Children in HH	0.035	(0.057)	-0.011	(0.060)
Year Dummies	Yes		Yes	
Region Dummies	Yes		Yes	
Observations	52,553		52,553	
Individuals	9,543		9,543	

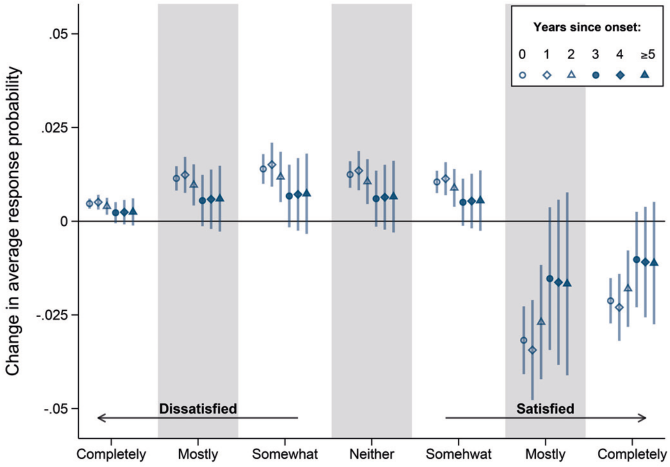
Source: Own calculations based on USoc Waves 2009-2020. *Note:* * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$). All analyses use cluster-robust standard errors at the individual level.

In ordered response models coefficient size provides limited information. While one can infer the initial differences in life satisfaction and self-assessed health to attenuate after LSI-onset, it is not possible to infer whether these changes are of similar magnitude or response probabilities are changed. To ease interpretation, we follow Baetschmann et al. (2020), by calculating the marginal effect on the average response probability with results depicted in Figure 2 for life satisfaction panel (a) and self-assessed health panel (b). The vertical y-axis indicates the change in the average response probability in percentage points with 0.05 indicating an increase of 5 percentage points. The horizontal x-axis shows the response categories for each outcome from lowest (left) to highest (right). Within response categories for each year spent living with an LSI the changes in the average response probabilities are plotted from onset (hollow dot, left) to five years and above (filled out triangle, right). Bars indicate 95-% confidence intervals. By construction, the sum of all changes must be zero as relative response probabilities simply shift along the underlying distribution with categories always summing up to one.

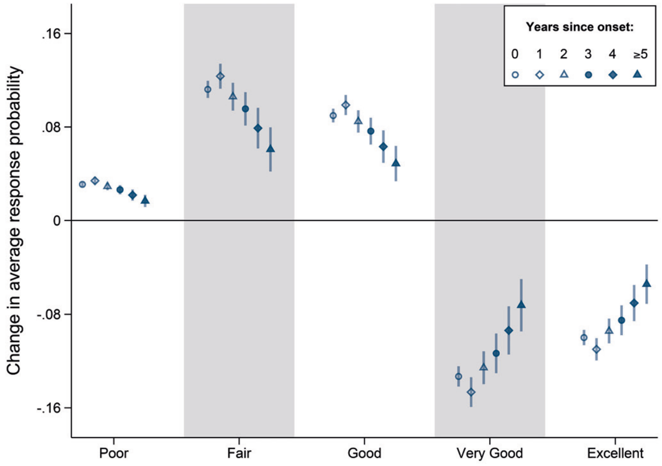
Figure 2 illustrates a general characteristic of our results. Life satisfaction and self-assessed health have a skewed distribution with most respondents reporting the highest two categories before LSI-onset (see Online Supplement Table O3.A1.1). This is reflected in the estimated response probability changes (Figure 2). Living with an LSI decreases the likelihood to respond in the categories (completely or mostly satisfied with one's life and very good or excellent health) while it increases the probability to respond in the remaining. Over time, these initial changes become smaller with average response probabilities trending towards levels observed before the onset for both measures, although at differential rates. For life satisfaction, the likelihood of responding with either of the highest categories drops by 5.3 percentage points. Given that about 59.2% of respondents reported these levels of life satisfaction before onset (see Online Supplement Table O3.A1.1) a decrease of around 9%. Among those living with an LSI for five or more years, this change in probability decreases to only -2.8 percentage points (4.6%) and becomes insignificant. For self-assessed health the picture is similar, but effect sizes are larger. Individuals are about 23 percentage points less likely to report very good or excellent health at the onset of an LSI, corresponding to a 43% reduction within the response group given the pre-onset share of 54% in these categories (see Online Supplement Table O3.A1.1). Five years after onset, the difference persists but decreases to only 12.6 percentage points (23%). For all subsequently presented results the response shares before onset of an LSI used to calculate relative changes listed in brackets can be found in Online Supplement Table O3.A1.1.

Figure 2: Marginal Effects by Outcome Variable - Full Sample

(a) Life Satisfaction



(b) Self-Assessed Health



Source: Own calculations based on USoc Waves 2009-2020. *Note:* Bars indicate 95-% confidence intervals.

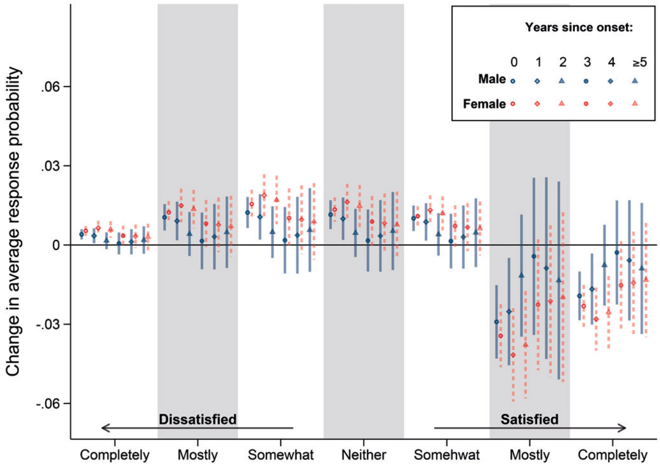
4.3 Results by Subgroups

To explore whether our results differ by subgroups, we divide our sample into male and female respondents. Figure 3 depicts the estimated changes in the marginal response probabilities for males (blue/solid) and females (red/dashed) for both outcomes (see Online Supplement Table O3.A2.1 for coefficient estimates). With respect to life satisfaction, LSI-onset is associated with a comparable decrease in the probability of responding in the highest two categories for both genders. For men the decrease is 4.8 percentage points (8.1%) and for women it is

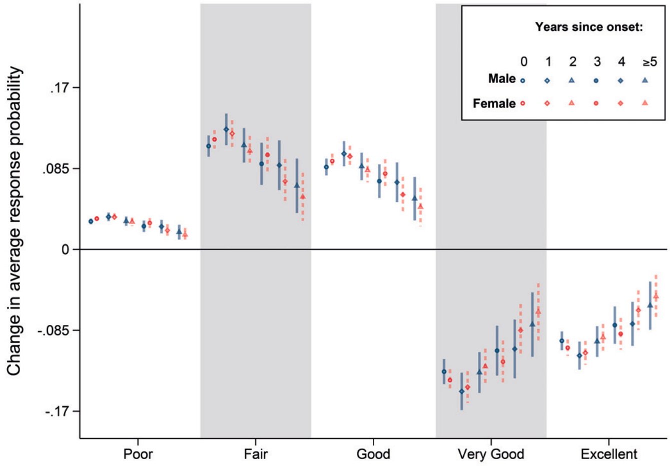
5.7 percentage points (9.6%). However, for men already after two years there is no longer a significant difference. For women, differences become insignificant after four years. With respect to self-assessed health the estimated changes in response probabilities are near identical across both groups.

Figure 3: Marginal Effects by Outcome Variable - Male vs Female Respondents

(a) Life Satisfaction



(b) Self-Assessed Health

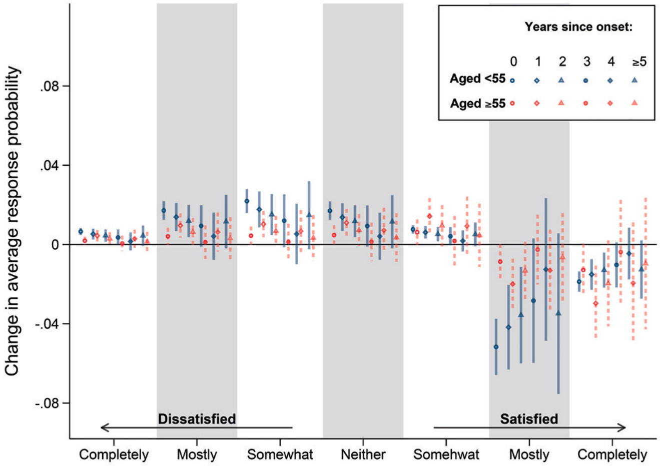


Source: Own calculations based on USoc Waves 2009-2020. *Note:* Bars indicate 95-% confidence intervals.

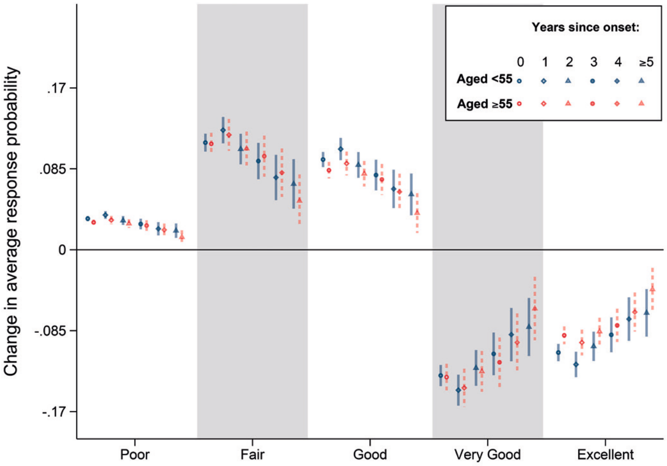
To explore whether the observed pattern of adaptation differs by age at onset, we divide our sample into two groups, those aged below 55 at LSI-onset and those older. This age threshold aligns with previous studies using datasets on older individuals aged 55 and older (Baji & Bíró, 2018; Bussière et al., 2021; de Hond et al., 2019). Figure 4 depicts the estimated changes in the response probabilities for individuals aged below 55 (blue/solid) and those aged 55 and above (red/dashed) for each outcome measure (see Online Supplement Table O3.A2.2 for coefficient estimates). For life satisfaction, we observe a clear difference between age groups. For younger individuals the onset of an LSI is associated with a stronger decrease in the likelihood of reporting high levels of life satisfaction. The onset is associated with a decrease of 7 percentage points (13.6%) while for older individuals this is only 2.2 percentage points (3.1%) and insignificant. For younger individuals, the initial changes in the average response probability decrease in size with time spent living with an LSI and become insignificant after three years. The point estimates for the average response probability changes for the lower life satisfaction categories are consistently higher for the younger age group than for the older respondents. Unlike for life satisfaction, we observe little difference between age groups for self-assessed health.

Figure 4: Marginal Effects by Outcome Variable - Younger vs Older Respondents

(a) Life Satisfaction



(b) Self-Assessed Health



Source: Own calculations based on USoc Waves 2009-2020. *Note:* Bars indicate 95-% confidence intervals.

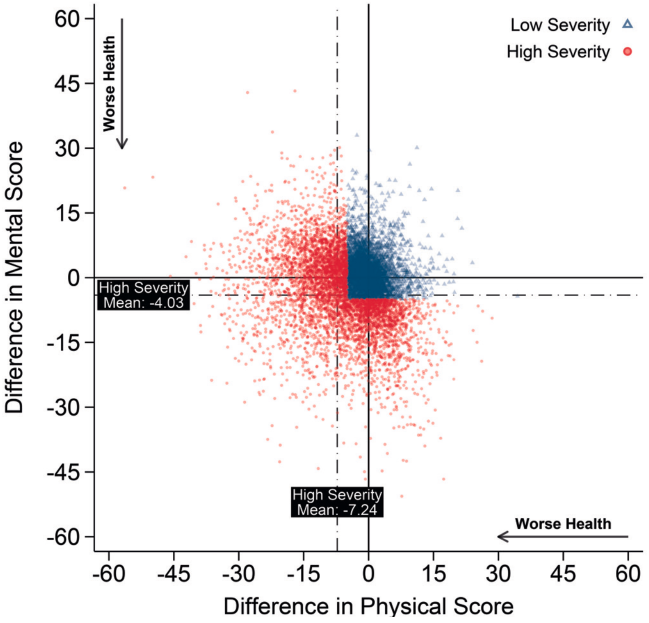
4.4 Severity of Health Changes and Subjective well-being

To explore whether adaptation is also observed across the intensity distribution with respect to health changes, we divide individuals into low and high severity groups based on observed differences in the SF12 component scores by computing the within-person difference between mental and physical health scores before and after LSI-onset. For the average person in the sample, the onset is associated with a deterioration in both dimensions, but for a considerable number of individuals the scores change only marginally. This is in line with the fact that only 54.49% of our analysis sample report a substantial limitation (see Online Supplement Table

O3.A1.2) and somewhat expected given the fuzzy definition of a long-standing illness in the survey as a mental or physical impairment, disability, or illness. We chose as our primary measure of severity the SF12 component score-changes as they provide a more granular measure of health changes. However, in the robustness checks we also consider alternative definitions of severity.

We categorize individuals into a high severity of health change group (red dots) if they report a difference of at least 5 points, or half a standard deviation, in either health dimension. For individuals in this high severity group (54% percent of the sample) changes are observed in both health dimensions, with the mean mental score dropping by 4 and the physical score by 7 points. The remaining respondents are included in the low severity group (transparent blue triangles). Figure 5 plots the individual-level differences for our analysis sample with mental health differences on the y-axis and physical health differences on the x-axis. Lower values indicate that after onset of an LSI the respective dimension score decreased.

Figure 5: Individual-Level Difference in Mean Mental/Physical Component Scores

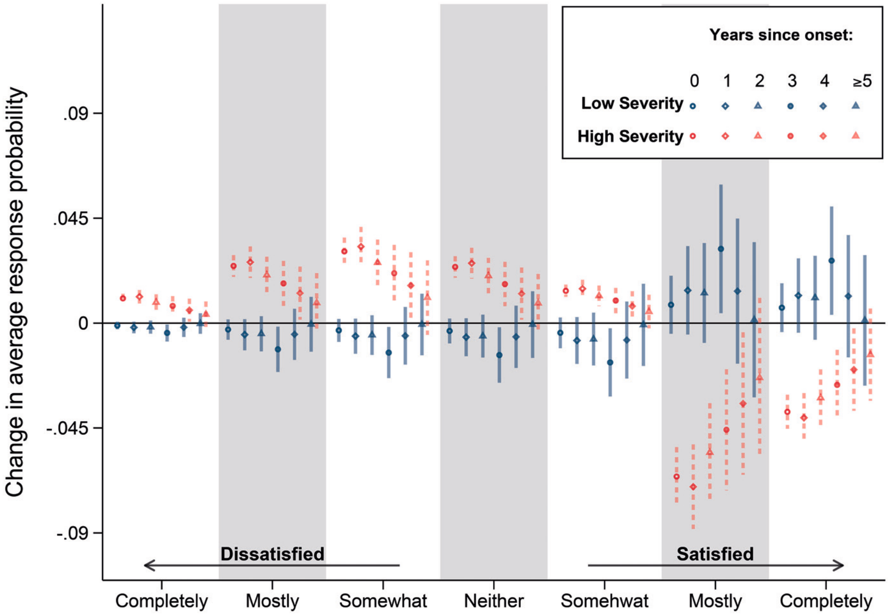


Source: Own calculations based on USoc Waves 2009-2020. *Note:* The difference is calculated by subtracting the individual-level post-onset mean of each score from the pre-onset mean.

This approach is used for exploring adaptation in life satisfaction but cannot be used for exploring adaptation in self-assessed health as mental and physical scores are based on all twelve items of the SF12 which include self-assessed health. Figure 6 depicts the estimated changes in the average response probability for life satisfaction comparing the low severity (blue/solid) with the high severity group (red/dashed) for both outcome measures. (see Online Supplement Table O3.A2.3 for coefficient estimates).

There is a clear relationship between the health change severity and the associated changes in response patterns. A more severe change in physical and/or mental health status is associated with a larger decrease in the predicted likelihood of reporting higher levels of life-satisfaction. In the high severity group, the initial onset is associated with a large decrease in the likelihood of reporting higher levels of life satisfaction. Individuals are 10.4 percentage points (18.6%) less likely to report to be completely or mostly satisfied with their lives. With increasing time spent living with an LSI, this initial difference decreases and becomes insignificant for the five years and above LSI-duration group. No significant changes in the average predicted response probability were observed in relation to the onset of an LSI of low severity.

Figure 6: Marginal Effects on Life Satisfaction by Severity



Source: Own calculations based on USoc Waves 2009-2020. *Note:* Bars indicate 95-% confidence intervals.

5 Robustness Checks

In the analyses so far, we follow recommendations by Frijters, Haisken-DeNew, and Shields (2004) only including year-dummies in combination with individual fixed-effects to capture ageing. This assumes ageing to have a homogeneous effect on our outcomes of interest across respondents. To explore whether this might be too restrictive we instead include cohort-year interaction terms to allow the effect of ageing to vary flexibly across 10-year birth cohorts. Online Supplement Figure O3.A3.1 depicts the resulting marginal effects (see Online Supplement Table O3.A3.1 for detailed results) comparing our baseline specification to the flexible specification allowing for age-specific effects. The results are nearly identical, confirming previously reported results by de Hond et al. (2019).

Further, to explore whether we observe patterns consistent with adaptation when changing the definitions of low and high severity groups we use a drop of at least 10 points (one standard deviation) in either the mental or physical health score as the cut-off. The results are depicted in Online Supplement Figure O3.A3.2 (detailed results are in Online Supplement Table O3.A3.2). The estimated changes in response probabilities are larger with this higher cut-off, but the overall picture is similar, showing that health changes of lower severity are associated with smaller changes in reported well-being. In a second step we consider severity levels defined by the number of areas of life dimensions reported to be substantially limited due to the LSI onset. This specification also allows us to explore differences in adaptation in self-assessed health between individuals reporting more severe health changes. Online Supplement Figure O3.A3.3 and Table O3.A3.3 compare our results for individuals in our sample reporting no dimension to be affected against those reporting at least one affected dimension. While more severe changes are associated with more pronounced changes in the response categories for both measures the overall pattern suggesting adaptation remains. A related concern follows from our reliance on the LSI question to identify transition into ill health, which is vague with respect to the severity of the underlying illness or disability and the onset-timing. As an alternative approach we use an item of the SF12 health questionnaire encoding whether respondents' health limits their daily activities. Please not that this results in an expanded dataset of different individuals as we now take the SF12-based limitation question and individuals' response patterns as the departure point to construct a dataset as done based on the LSI question. If a limitation is reported individuals indicate its severity as low (*"limited a little"*) or high *"limited a lot"*). Online Supplement Figure O3.A3.4 and Table O3.A3.4 depict the results of this alternative definition and suggest that adaptation occurs also when considering health changes associated with an actual functional limitation. As Online Supplement O3.A3.4 and Table O3.A3.4 show, the pattern for self-assessed health is highly similar to the pattern for life satisfaction. For both the observed patterns indicate that over time response probabilities trend towards their pre-onset levels while the severity of limitations does strongly correlate with the initial decrease.

A last concern relates to the role of disease duration, severity, and survey attrition. Individuals suffering from a severe health shock might be more likely to drop out of the panel over time.

If so, we would more observe individuals reporting an LSI at the lower end of the severity distribution, to which they may more easily adapt. To explore whether this is indeed the case, we repeat our baseline analysis on a separate sample of individuals that can be observed for at least four years after reporting an LSI. Online Supplement Figure O3.A3.5 and Table O3.A3.5 compare the results from the baseline sample to the sample with limited attrition for both outcomes. With respect to life satisfaction, the observed results are highly comparable although it seems that the response probabilities in the group with limited attrition return to levels before onset of the LSI more distinctly. For self-assessed health, the picture is less clear as we observe a higher decrease in the probability of reporting excellent health among the group with limited attrition, while also observing a smaller increase in the likelihood to report good health. On the other hand, we observe a slightly larger decrease in the probability of reporting their health to be very good and a larger increase in the likelihood to report fair or even poor health. However, as we still observe a pattern of attenuating changes across response categories over time, we see little conclusive evidence that the results found by our main analysis are driven by selective attrition alone.

6 Discussion

The propensity to adapt to deteriorating health is a desirable manifestation of psychological resilience at the individual level but may have undesirable implications in the context of health economic evaluations. If adaptation occurs, should it be considered when healthcare resources are allocated? What are the consequences of doing so when adaptation varies across types of conditions and subgroups? As increasingly outcome measures on subjective well-being and health are used to quantify the impact of ill health on individuals' quality of life the possibility of adaptation remains a concern, but empirical evidence is inconclusive. This study uses a general population survey from the UK and fixed effects ordered response models to explore whether and how people adapt to ill health. To do so we use the onset of a long-standing illness or disability in combination with data from SF12 health questionnaires to explore the extent of adaptation across in the domains of life satisfaction and self-assessed health.

Our analyses using life satisfaction as outcome of interest provide evidence that adaptation in well-being measures is substantial and occurs already after a short period. The onset of a long-standing illness or disability decreases the likelihood to report higher levels of life satisfaction considerably, but as time progresses individuals revert towards their levels of life satisfaction before onset after three years, that is, they adapt to persisting ill health. Further, we find evidence that the observed patterns differ across subgroups. Men adapt slightly quicker than women and for individuals younger than 55 at onset the impact of ill health is larger, and adaptation takes longer than for older individuals. Lastly, even in the case of more severe health changes and functional limitations we find adaptation to occur but taking a longer period.

When considering the impact of long-standing illness or disability on self-assessed health our results indicate a reverting trend towards pre-onset levels of self-assessed health before the onset of a long- standing illness but no full adaptation. The onset of ill health results in a large decrease in the likelihood of reporting high levels of subjective health. This effect persists even after five years but the magnitude of these differences decreases. The indicates that while individuals adapt to ill health, it is not sufficient to offset the full impact over time. Interestingly, we do not observe significant differences in adaptation across subgroups.

Overall, our results confirm findings from de Hond et al. (2019) in a sample of individuals aged 55 and older. In addition, our results across specifications are supportive of Powdthavee (2009) and suggest that adaptation in life satisfaction coincides with adaptation in self-assessed health. However, we are also able to provide some complementary perspectives. Adaptation in life satisfaction differs considerably by age while this is not the case for self-assessed health. This could indicate that as individuals age the importance of health for well-being diminishes (Bussière et al., 2021; Frijters, 2000) or that older individuals are more resilient to such shocks (Etilé, Frijters, Johnston, & Shields, 2021), potentially because deteriorating health is more expected and accepted, and more common among these age-groups. Our results by gender provide a similar picture with respect to differences being mainly observed when considering life satisfaction but not self-assessed health as the outcome of interest, although the differences between groups are less pronounced than in case of the age-groups.

The results of this study therefore provide useful insights to policymakers and researchers interested in measuring health-related quality of life changes using subjective outcome measures. Exploring these effects, including the role of adaptation, requires routinely collecting data on disease onset and duration as well as measures of outcomes relevant to individuals and appropriate for policy evaluation. This is especially relevant for studies relying on empirical approaches that exploit the availability of longitudinal data. For example, an increasing literature has applied well-being valuation approaches to explore the monetary equivalent of specific health conditions (see for example recently by Howley (2017) or Ólafsdóttir et al. (2020)) or summary measures such as quality-adjusted life years (Himmler et al., 2021; Huang et al., 2018). Adaptation decreases the monetary value of health obtained from such studies. Considering the possibility of adaptation should therefore be taken into consideration when interpreting their results, for example in health economic evaluations.

6.1 Limitations and Future Research

Our study has several limitations. We use an ordered response model for life satisfaction and self-assessed health, which differs from most previous studies on the topic of adaptation that use an ordinary least squares fixed-effects design. While our approach provides additional insights it does not alter the underlying assumptions. For a causal interpretation of our results the time-invariant unobservable differences captured by the fixed-effects approach need to be the only source of unobserved confounding. Assuming that the onset of an LSI is exogenous conditional on such characteristics is more credible than without addressing time-invariant unobservable differences, but it remains a strong assumption. Further, in our analyses we rely on a broad classification of ill health using LSI-onset and reported functional limitations while we consider different definitions of periods of ill health based on other measures. This approach has the benefit of relying on individuals that report an ongoing change to their health as opposed to a diagnosis of a specific disease of which the impact on patients' lives is not always consistent over time (de Hond et al., 2019). This is also consistent with the emphasis in the health economic evaluation literature, which predominantly considers measures such as the EQ-5D, to quantify the burden from diseases, given their impact on different dimensions of health-related quality of life irrespective of the specific diagnosis. At the same time, the underlying health issues or diseases causing reported health changes to remain unobserved but could be interesting with respect to the heterogeneous impact of different diseases (Binder & Coad, 2013; Graham et al., 2011). It further raises the questions on the role of health events as sources of information shaping perceived health (Nielsen, 2016) and health behaviors (see Gaggero et al., (2022) and Verdun (2022)). Previous studies such as Cubí-Mollá et al. (2017) and Baji and Biro (2018) have tried to explore jointly whether specific diagnoses underlying the reported LSI result in different patterns of adaptation. However, while general population surveys as used in this study allow for groups of individuals to be observed the number of individuals transitioning into specific conditions is limited. Lastly, we have to rely overall on self-reported information and cannot use objective health measures. Further, while the USoc surveys cover an extensive set of topics they provide only limited information on what type of medical care individuals use. Therefore, we cannot distinguish whether the observed adaptation pattern could partially be explained by recovery, treatment success or the use of (medical) devices that help individuals to function and participate in activities of daily life despite functional limitations. This is an important caveat deserving emphasis because it indicates that the degree to which the adaptation commonly found in empirical studies for certain health changes may in part be a natural occurrence as patients receive treatment or manage their symptoms. This raises important questions on how adaptation should be considered in the context of health economic evaluations, given the ethical dilemmas this implies (Brazier et al., 2018).

The limitations of this study do provide guidance on future research directions. One way to address many of the discussed concerns would be to combine survey data with administrative data that would allow researchers to rule out or at least quantify any bias resulting from unobserved information. Such data could for example be used to determine whether individuals generally participating in such surveys and suffering from a specific disease are overall representative of the general patient population they belong to. Detailed information on hospital admissions and past or future healthcare use would also allow researchers to cross-validate the self-reported information. Such information would also allow a relaxation of the main identifying assumption. While a causal interpretation requires us to assume the health shock (LSI-onset) to be exogenous administrative records would allow one to employ a different identification strategy. Instead of relying on the shock itself to be exogenous one could exploit the exogenous timing of certain health shocks, such as heart attacks or strokes (Fadlon & Nielsen, 2019), to compare individuals that have suffered from such a shock to those that will suffer from it soon. Such an approach might also allow for a more direct disentangling of the different effects a diagnosis might have on subjective well-being and health outcomes. Focusing on the onset of conditions that have strong implications for individuals past and future health would allow to identify whether observed changes are driven by realised health changes i.e., the loss of certain functionalities, as opposed to the information gain resulting from a diagnosis or health event.

7 Conclusion

In summary, our study provides new evidence on individuals' propensity to adapt to health across measures of subjective health and well-being. We observe adaptation to occur in both domains while broader subjective well-being measures such as life satisfaction seem to be more strongly affected by this phenomenon. Our results also indicate that adaptation is not limited to only the mildest health shocks and that it varies across certain subgroups such as gender and age. While from the individual's perspective adaptation is desirable, it poses problems for the application of subjective outcome measures in health economic evaluations. Adaptation should therefore remain a concern for researchers aiming to complement domain specific, objective quality of life measures with broader subjective outcome measures. While the contemporary impact of ill health is captured by such measures, adaptation can be significant even in the case of severe health shocks that lead to functional limitations. As such adaptation remains an obstacle in research practice, and the benefits of using subjective outcome measures should be weighed against the drawbacks resulting from adaptation when applied in practice.

References

Arni, P., Dragone, D., Goette, L., & Ziebarth, N. R. (2021). Biased health perceptions and risky health behaviors—theory and evidence. *Journal of Health Economics*, 76, 102425.

Baetschmann, G., Ballantyne, A., Staub, K. E., & Winkelmann, R. (2020). feologit: A new command for fitting fixed-effects ordered logit models. *The Stata Journal*, 20(2), 253--275.

Baetschmann, G., Staub, K. E., & Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), 178(3), 685--703.

Baji, P., & Bíró, A. (2018). Adaptation or recovery after health shocks? Evidence using subjective and objective health measures. *Health Economics*, 27(5), 850-864.

Benjamin, D. J., Cooper, K. B., Heffetz, O., & Kimball, M. S. (2019). Self-reported wellbeing indicators are a valuable complement to traditional economic indicators but aren't yet ready to compete with them. *Behavioural Public Policy*, Forthcoming.

Binder, M., & Coad, A. (2013). "I'm afraid i have bad news for you…" estimating the impact of different health impairments on subjective well-being. *Social Science & Medicine*, 87, 155--167.

Bond, T. N., & Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, 127(4), 1629--1640.

Brazier, J., Rowen, D., Karimi, M., Peasgood, T., Tsuchiya, A., & Ratcliffe, J. (2018). Experience-based utility and own health state valuation for a health state classification system: why and how to do it. *The European Journal of Health Economics*, 19(6), 881--891.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 36(8), 917.

Bussière, C., Sirven, N., & Tessier, P. (2021). Does ageing alter the contribution of health to subjective well-being? *Social Science & Medicine*, 268, 113456.

Chen, L.-Y., Oparina, E., Powdthavee, N., & Srisuma, S. (2022). Robust ranking of happiness outcomes: a median regression perspective. *Journal of Economic Behavior & Organization*, 200, 672--686.

Cohen, G. A. (1993). Equality of what? on welfare, goods, and capabilities. *The Quality of Life*, 9--29.

Cubí-Mollá, P., Jofre-Bonet, M., & Serra-Sastre, V. (2017). Adaptation to health states: Sick yet better off?. *Health Economics*, 26(12), 1826-1843.

de Hond, A., Bakx, P., & Versteegh, M. (2019). Can time heal all wounds? an empirical assessment of adaptation to functional limitations in an older population. *Social Science & Medicine*, 222, 180--187.

Dickerson, A., Hole, A. R., & Munford, L. A. (2014). The relationship between well-being and commuting revisited: Does the choice of methodology matter? *Regional Science and Urban Economics*, 49, 321--329.

Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *The Economic Journal*, 118(525), 215--234.

Dolan, P., & Metcalfe, R. (2012). Valuing health: a brief report on subjective well-being versus preferences. *Medical Decision Making*, 32(4), 578--582.

Etilé, F., Frijters, P., Johnston, D. W., & Shields, M. A. (2021). Measuring resilience to major life events. *Journal of Economic Behavior & Organization*, 191, 598--619.

Fadlon, I., & Nielsen, T. H. (2019). Family health behaviors. *American Economic Review*, 109(9), 3162--91.

Ferrer-i Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497), 641--659.

Frederick, S., & Loewenstein, G. (1999). 16 hedonic adaptation. Well-Being. The foundations of Hedonic Psychology/Eds. D. Kahneman, E. Diener, N. Schwarz. NY: Russell Sage, 302--329.

Frijters, P. (2000). Do individuals try to maximize general satisfaction? *Journal of Economic Psychology*, 21(3), 281--304.

Frijters, P., Clark, A. E., Krekel, C., & Layard, R. (2020). A happy choice: wellbeing as the goal of government. *Behavioural Public Policy*, 4(2), 126--165.

Frijters, P., Haisken-DeNew, J. P., & Shields, M. A. (2004). Money does matter! evidence from increasing real income and life satisfaction in east germany following reunification. *American Economic Review*, 94(3), 730--740.

Gaggero, A., Gil, J., Jiménez-Rubio, D., & Zucchelli, E. (2022). Does health information affect lifestyle behaviours? The impact of a diabetes diagnosis. *Social Science & Medicine*, 314, 115420.

Gan, L., Gong, G., Hurd, M., & McFadden, D. (2015). Subjective mortality risk and bequests. *Journal of Econometrics*, 188(2), 514--525.

Graham, C., Higuera, L., & Lora, E. (2011). Which health conditions cause the most unhappiness? *Health Economics*, 20(12), 1431--1447.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, 19(3), 403--420.

Hagenaars, A. J., De Vos, K., Asghar Zaidi, M., et al. (1994). Poverty statistics in the late 1980s: Research based on micro-data. Office for Official Publications of the European Communities.

Himmler, S., Stöckel, J., van Exel, J., & Brouwer, W. B. (2021). The value of health—Empirical issues when estimating the monetary value of a quality-adjusted life year based on well-being data. *Health Economics*, 30(8), 1849-1870.

Howley, P. (2017). Less money or better health? evaluating individual's willingness to make trade-offs using life satisfaction data. *Journal of Economic Behavior & Organization*, 135, 53--65.

Huang, L., Frijters, P., Dalziel, K., & Clarke, P. (2018). Life satisfaction, qalys, and the monetary value of health. *Social Science & Medicine*, 211, 131--136.

Hünermund, P., & Louw, B. (2020). On the nuisance of control variables in regression analysis. arXiv preprint arXiv:2005.10314.

Jonker, M. F., Attema, A. E., Donkers, B., Stolk, E. A., & Versteegh, M. M. (2017). Are health state valuations from the general public biased? a test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. *Health Economics*, 26(12), 1534--1547.

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences,* 119(42), e2210412119.

Kaiser, C., Vendrik, M., et al. (2019). How threatening are transformations of reported happiness to subjective wellbeing research? (Tech. Rep.). Center for Open Science.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2), 391--413.

Loewenstein, G., & Ubel, P. A. (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*, 92(8-9), 1795--1810.

Lucas, R. E. (2007). Adaptation and the set-point model of subjective well-being: Does happiness change after major life events? *Current Directions in Psychological Science*, 16(2), 75--79.

McNamee, P., & Mendolia, S. (2014). The effect of chronic pain on life satisfaction: evidence from australian data. *Social Science & Medicine*, 121, 65--73.

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social Science & Medicine*, 55(12), 2149--2158.

Muris, C. (2017). Estimation in the fixed-effects ordered logit model. *Review of Economics and Statistics*, 99(3), 465--477.

Nielsen, T. H. (2016). The relationship between self-rated health and hospital records. *Health Economics*, 25(4), 497-512.

Ólafsdóttir, T., Ásgeirsdóttir, T. L., & Norton, E. C. (2020). Valuing pain using the subjective well-being method. *Economics & Human Biology*, 37, 100827.

Oswald, A. J., & Powdthavee, N. (2008). Does happiness adapt? a longitudinal study of disability with implications for economists and judges. *Journal of Public Economics*, 92(5-6), 1061--1077.

Peasgood, T., Foster, D., & Dolan, P. (2019). Priority setting in healthcare through the lens of happiness. Global Happiness and Well-Being Policy Report.

Powdthavee, N. (2009). What happens to people before and after disability? focusing effects, lead effects, and adaptation in different areas of life. *Social Science & Medicine*, 69(12), 1834--1844.

Schünemann, J., Strulik, H., & Trimborn, T. (2017). Going from bad to worse: Adaptation to poor health health spending, longevity, and the value of life. *Journal of Economic Behavior & Organization*, 140, 130--146.

Spaenjers, C., & Spira, S. M. (2015). Subjective life horizon and portfolio choice. *Journal of Economic Behavior & Organization*, 116, 94--106.

Spitzer, S., & Shaikh, M. (2022). Health misperception and healthcare utilisation among older europeans. *The Journal of the Economics of Ageing*, 22, 100383.

Ta, A. (2019). Adaptation to disability-evidence from the UK household longitudinal study. *The Sheffield Economic Research Paper Series* (SERPS), 2019020(2019020).

University of Essex Institute for Social and Economic Research. (2022). Understanding society: Waves 1-12, 2009-2022 and harmonised bhps: Waves 1-18, 1991-2009. [data collection] 17th edition. UK Data Service. SN: 6614.

Verdun, Z. S. (2022). Disease Realization Versus Risk Warning: Impact on Lifestyle Behaviours. Working Paper

Versteegh, M., & Brouwer, W. (2016). Patient and general public preferences for health states: a call to reconsider current guidelines. *Social Science & Medicine*, 165, 66--74.

Ware, J. E., Keller, S. D., & Kosinski, M. (1995). SF-12: How to score the sf-12 physical and mental health summary scales. Health Institute, New England Medical Center.

Westreich, D., & Greenland, S. (2013). The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology*, 177(4), 292—298.

4

CHAPTER 4

Staying fit but feeling better? – The Impact of Health Shocks on Health Perceptions and Behaviours

with Pieter Bakx and Bram Wouterse

Abstract:

We study health events as sources of information shaping health perceptions and behaviours within families. By combining individual-level administrative data and survey data from the Netherlands we estimate the short-, medium- and long-term causal effect of suffering from a heart attack or stroke on the health perceptions and behaviours of patients and spouses of patients. Our identification strategy exploits the exogenous timing of these health shocks in a treated vs not-yet-treated event-study design. Experiencing a heart attack or stroke leads to a strong downward adjustment in perceived health remaining persistent for multiple years after the health event. This adjustment persistent when accounting for health information individuals have at the time of the health event and goes beyond the realized changes in latent health status, measured using functional limitations, suggesting long-term persistent learning effects. In line with this, we observe substantial decreases in smoking prevalence and alcohol consumption and an uptake of preventive medication use. Unlike the strong effects across domains found among patients themselves, we find only limited evidence for spillover effects on spouses of patients after controlling for changes in their latent health. Our results underline the importance of experienced health events as drivers of changes in health perceptions and behaviours.

1 Introduction

The ability to accurately process information about one's health is a prerequisite for optimal choices e.g., regarding life-cycle planning or investments in health prevention (Grossmann, 1972), and a key assumption underlying many health-related policy reforms. Severe shocks such as heart attacks or strokes are salient signals providing information on individual-level risk factors with long-tern implications for survival and hence for life-cycle planning and behaviour change. Salient cues can be a trigger for individual-level behaviour change (Bernheim & Rangel, 2004) but the persistence of these changes can be undermined if salience decreases with increasing time distance to the event, for example due to inattention (Bordalo et al., 2009; DellaVigna, 2009). Whether the experience of health events lead to sustained behaviour change with respect to amenable risk factors therefore rests on whether perceptions about one's health are updated in the right direction but also persistently.

The formation of individual-level health perceptions is inherently difficult given the uncertain nature of health outcomes over the life-course (Cropper, 1972). While across diverse populations measures of perceived health are found to be highly correlated with objective health dimensions there is considerable heterogeneity in the relationship between specific dimensions of objective health information and perceived health and with respect to the strength of these relationship across subgroups (see e.g., Jürges (2007)). A particular concern is the existence of health perception biases, the systemic over- or under-estimating of one's health conditional on objective health information. Biased health perceptions have been linked to increased engagement in certain risky behaviours (Arni et al., 2021; Nie et al., 2021), decreased health prevention efforts (Bíró, 2016; Spitzer & Shaikh, 2022), and could result in suboptimal life-cycle planning (Gan et al., 2015; Bago d'Uva, 2019). In the context of health shocks this discrepancy between objective health changes and perceived health emerging over time is often referred to as adaptation (see for example Groot (2002) or Menzel et al. (2002) Groot (2000)). Adaptation is of particular concern in the case of informative health events. Past studies suggest that even severe health shocks result only in a transient adjustment to subjective survival probabilities that do not reflect persistent changes in objective health with implications for long-term persistent behaviour changes (Baji & Biro, 2018).

The role of health events as information shocks is also relevant due to the potential for learning effects expanding within and between households. Especially with respect to spouses, a large body of literature has documented high levels of concordance in the health of couples (Meyler, 2007). While these correlations vary across countries the overarching relationship is highly persistent (Banks et al., 2021) and reflected in the spousal correlations with respect to risky health behaviours and prevention efforts (see for example Cutler & Lleras-Muney (2010), Cawley & Ruhm (2011), and recently Bouckaert et al. (2020)). Health shocks therefore often have informational value not only for patients. In line with this Fadlon & Nielsen (2019) have documented causal evidence for far reaching behavioural spillovers from cardiovascular shocks on health prevention efforts, especially statin use, of spouses but also adult children and even colleagues of patients. Whether these shocks also affect how individuals perceive their own health, potentially indicating learning effects from observed health events, is less clear but relevant for a better understanding of the behavioural origins of spillover effects.

This paper studies how health events shape health perceptions and behaviours of patients and family members of patients, namely spouses. We estimate the long-term causal effects of experiencing health events directly (patients) or indirectly by observing the occurrence of such events in one's social network (spouses of patients) on health perceptions and behaviours to quantify the learning effects resulting from such events. Our identification strategy exploits the exogenous timing of heart attacks and strokes in an event-study design in which we compare individuals exposed to the same shock but at different time-points and who shared the same *ex-ante* risk of a direct or indirect exposure (see also Fadlon & Nielsen (2019) or Chandra and Staiger (2007) who use a similar identification strategy). To do so we exploit a unique combination of survey and administrative data. We combine three waves of the *Dutch Health Monitor* survey with longitudinal administrative data on the Dutch population on healthcare use (inpatient hospitalizations and prescription medication use), socio-economic and demographic status, and death records. This combination allows us to link how individuals perceive their own health to be with different dimensions of objective health to quantify the magnitude and persistence of learning effects resulting from heart attacks and strokes.

We first consider how non-fatal health events serve as information shocks shaping health perceptions of patients and spouses of patients over the long-term. To do so we document the impact of these events on self-perceived health conditional on objective health information at the time of the event and different dimensions of health-related functioning after the health shock. Subsequently we turn towards their causal effect on risky and preventive health behaviours. We focus on smoking, alcohol consumption, exercising habits and weight management due to their role as significant but avoidable risk factors for heart attacks and strokes (Feigin et al., 2017; Roth et al., 2020). Further we consider adherence to prescription medications as a crucial preventive health behaviour among populations with increased risk for cardio-vascular events (Binno, 2016) whose effectiveness requires long-term adherence (Choudry et al, 2011). The combination of self-reported data on health behavioural risk factors alongside administrative data on individual level healthcare use enables us to study the causal effect of health events on various health behaviours and is a key advantage of our study's setting.

This paper makes two contributions to the existing literature on how experienced health changes impact health perceptions and behaviours. Firstly, due to our unique setting allowing us to exploit the exogenous timing of severe and highly salient health events we provide a clean identification of the long-term effects of experiencing health shocks on individuals' health perceptions. By combining information on different dimensions of objective and subjective health around the onset of these health shocks we estimate whether health events result in long-term persistent adjustments to perceived health conditional on objective health information, suggesting learning effects, or whether perceptions are only updated temporarily due to a salient information shock. Secondly, we also identify the causal effect of these health events on health-related behaviours indicating increased individual-level investments in health prevention across different dimensions, and whether behavioural adjustments persist over time. This allows us to provide evidence on how experienced health events shape health perceptions but also whether these adjusted perceptions have economically relevant implications for health-related behaviours. Our results therefore provide insights into health events as factors shaping health perceptions and behaviours over the life-course through learning effects.

We further contribute to the literature on spillover effects and social learning from health events. There is a growing body of literature documenting how the impact of health shocks on individual-level behaviours propagates across social networks. Our setting allows us to explore spillover effects from health events across a range of self-reported health-related behaviours alongside preventive healthcare consumption, therefore providing evidence across different dimensions of health-related behaviours. More importantly, we can explore whether a partner's health event directly alters their spouse's health beliefs conditional on pre-existing health information before the health event and current objective health. The few studies that have considered the role of health shocks observed within the family on both health perceptions and behaviours provide mostly descriptive evidence directly identifying such learning effects (see e.g., Darden & Gilleskie, 2016). We therefore provide novel evidence on both the existence and magnitude of behavioural spillovers and whether these behavioural adjustments are in line with estimated learning effects in how individuals perceive their own health.

Our findings on the causal effects of health events on patient's health perceptions indicate substantial and long-term persistent learning effects. We first illustrate how experiencing a heart attack or stroke results in substantial adjustments to perceived health and that these adjustments are highly persistent with only minor attenuation over time. Using a combination of information on objective health information on the *ex-ante* health risk individuals had at the time of the event and their latent health across dimensions of functional limitations we further show that changes to perceived health go beyond objective health changes. While both heart attacks and strokes have a negative effect of the prevalence of physical and cognitive/sensory functional limitations the changes in perceived health persist when accounting for objective health differences and pre-existing health information. This implies that changes to individual's health perceptions following health events incorporate new information and adjustments reflect learning effects about one's health that persist for up to seven years after the event.

Consistent with health events affecting health perceptions of patients through persistent learning effects we document substantial and persistent behavioural adjustments indicating increased individual-level investments in health prevention across multiple dimensions. We first document the causal effects of heart attacks and strokes on risky health behaviours in the form of smoking and alcohol consumption. With respect to smoking behaviour both events have economically relevant and highly persistent effects. Heart attacks result in 14 percentage point decrease in the smoking rate (ca. 46% against the baseline smoking rate) while strokes result in a 10 percentage point decrease (ca. 40% against the baseline smoking rate). The impact on alcohol consumption is smaller but still relevant with the number of drinks a week decreasing by an equivalent of 25% compared to the baseline. With respect to exercising frequency and bodyweight we find no evidence of either health event to result in behavioural adjustments. However, when considering the uptake and adherence of different types of preventive medications we observe highly persistent changes in medication use. Across all three considered types of prescription medications, anti-thrombotic, anti-hypertensive, and lipid-lowering (statins), we find heart attacks and strokes to result in near-universal uptake and persistent adherence over time.

While we find health events to result in long-term persistent learning effects and corresponding investments in preventive health behaviours for patients, we find little evidence for either of these with respect to their spillover effects on spouses. In line with our approach for patients we first document the effect of a partner's health event on their spouse's perceived and objective health and find no evidence for learning effects. Turning to spousal investments in preventive health behaviours, we also find only scarce evidence for behavioural adjustments. Neither for the four considered dimensions of self-reported health behaviours nor with respect to prescription medication use we find evidence for a partner's health event to result in spouses decreasing their engagement in risky health behaviours or increasing their health prevention efforts. However, while taken at face value these results provide consistent evidence against behavioural spillovers on spouses, we also provide evidence that at least in part these results are driven by methodological choices.

The remaining sections of this paper are organized as follows. Section 2 provides an overview of the data sources used. Our empirical strategy is discussed in Section 3 alongside the interpretation of our empirical estimates based on a conceptual model motivating the general methodological approach. Section 4 presents the results, starting with the causal effects of health events on the perceptions and behaviours of patients before considering spillovers on the souses of patients. Robustness checks are considered in Section 5 before concluding this paper in Section 6 with a discussion of our results and their implications for future research and policy.

2 Data

2.1 Data Sources

We use the general population samples of the Dutch Health Monitor (*Gezondheidsmonitor*) of 2012, 2016 and 2020. The health monitor is a large-scale, nationally representative survey of the general adult population in the Netherlands (aged 18+) aimed at measuring population health and organized by Statistics Netherlands (*Centraal Bureau voor de Statistiek*, CBS). Starting in 2012 it is conducted every four years. Selected individuals registered in their respective municipality are invited to participate with the aim of obtaining a representative sample of the adult population of the Netherlands. Only those individuals living in an institutionalized setting, such as permanent nursing home residents, are not approached. The collection of surveys is conducted by municipalities each year within the same period of September 1st to December 31st by inviting individuals by letter to participate in the online survey. A paper-based questionnaire was included in some regions in the invitation letter while in others this was only sent to those not responding to the initial invitation. Most responses were submitted online with in-person or phone-based interviews making up only 0.5% (2012) and 0.1% (2016/2020) of collected responses (CBS, 2015; 2017). The 2012 Health Monitor was sent to approximately 700,000 individuals with 387,195 respondents (ca. 55%), while in 2016 1.15 million individuals were approached resulting in 457,153 respondents (ca. 40%) followed by 1.39 million individuals yielding 539,000 respondents (ca. 39%) in 2020.

All health monitor surveys are linked by Statistics Netherlands to Dutch administrative records using pseudo-anonymized unique individual-level identifiers. This administrative data covers multiple dimensions; demographic background information (age, gender, and, if applicable, time and cause of death), socio-economic variables (household income based on income tax data) and healthcare use (hospitalizations and medicine consumption). Table O4.A1.1 in the Online Supplement provides an overview of the included data sources. The administrative data on medicine consumption we include in our analysis covers the period of 2006 to 2020. As parts of our analyses use lagged information on these indicators for up to three years before a health event this results in an effective observation period ranging from 2009 to 2020.

2.2 Outcome Measures: Health Perceptions and Behaviours

Our main outcome is an indicator of perceived health measured using general self-assessed health. Self-assessed health is measured on a five-point likert scale ranging from best health (1: *Very Good*) to worst possible health (5: *Very Poor*). This represents one of the broadest conceptualizations of subjective or perceived health. It is a heavily used measure in empirical research, often as an outcome to capture health effects directly, as it represents a composite measure combining a range of different health dimensions (Au & Johnston, 2014). It has been shown that self-assessed is predictive of future health outcomes even after accounting for objective health information, suggesting this measure to contain substantial private information of respondents on their own health state (Nielsen, 2016). Ultimately, however, it is a subjective measure with a range of response biases known to occur with a long-standing but ongoing discussion of its reliability (see for example Lindeboom & van Doorslaer (2004) or Arni et al. (2021) for a discussion). Further, especially in the context of health events studies suggest that self-reported health, or even specific measures like subjective survival probability, are only temporarily affected by health shocks even in the case of severe events like cancer diagnoses (Baji & Biro, 2018). Studying the effect of health shocks on self-perceived health as a measure of health perceptions is therefore motivated both by the measure's economic relevance and the potential to quantify the magnitude and persistence of learning effects from health events.

Our analysis of health behaviours uses self-reported information collected as part of the health monitor surveys. Smoking status is based on individuals reporting to be smokers at the time of their survey response without making a distinction on the intensity of their smoking habit. Alcohol consumption is based on the self-reported number of drinks consumed in an average week. Self-reported data on individuals' height and weight is used to compute the body-mass index while physical activity is self-reported as days in each week in which at least 30-60 minutes of moderate to strenuous physical activity are conducted. All these behaviours are chosen as they play a key role in decreasing overall risk of cardiovascular disease and to decrease the risk of subsequent heart attacks and strokes. Smoking and excessive alcohol consumption are important individual-level modifiable risk factors improving post-shock outcomes and singled out as key priorities for policy interventions. (Feigin et al., 2017; Roth et al., 2020).

Smoking cessation, alcohol consumption and overall physical fitness are important behavioural changes individuals can enact to improve their health outcomes and decrease cardio-vascular risk. Another important factor is the adherence to prescription medication regiments. Pharmaceutical innovations have been found to be a crucial driver of the increase in life expectancy in the US and with respect to cardio-vascular health (Buxbaum et al., 2020). In Europe clinical guidelines recommend the sustained long-term use of anti-hypertensive/anti-thrombotic medications and statins to prevent subsequent heart attacks and strokes (Binno, 2016) and improve long-term survival. Adherence to long-term therapy is however often only partial with patients discontinuing therapy. This has been a persistent finding in the literature even when consumption is not associated with direct healthcare costs to the individual (Choudry et al, 2011). While adjustments in smoking behaviour, alcohol consumption and physical fitness overall might have tangible short-term benefits to the individual adherence to medication therapy might not, due to its preventive nature with few directly observable health improvements. As such, especially for medication adherence, health perceptions might play a crucial role in determining behaviour as the consequences of non-adherence are not directly obvious.

2.3 Measure of objective health

As our primary measure of objective health, we use self-reported functional limitations which can be divided into limitations to physical functioning and cognitive or sensory functioning. The five cognitive or sensory functions covered are not being able to follow a conversation of three or more persons, not being able to have a conversation with one person, not being able to read the small print in newspapers, not being able to recognize someone's face at a distance of four meters or more. The three domains of physical functioning are not being able to carry 5kg for 10 meters; not being able to reach the ground; not being able to walk 400 meters without stopping. Each of these dimensions is surveyed using a question on the degree to which the activity described can be performed by the individual. The response options are whether an activity can be done without any limitation, with some effort, with high effort, or not at all. These measures are chosen as they are the most detailed available measure on health-related functionality across different domains in all three health monitors. A focus on these measures is also in line with other studies that primarily considered how perceived health maps on to more objective measures of latent health such as Dwyer and Mitchell (1999), Jürges (2007), or more recently Au and Johnston (2014). Nonetheless we are aware that the focus on functional limitations runs the risk of omitting other dimensions of objective health, especially mental health or other dimensions not represented in current latent health but rather the overall risk individuals have for certain health events.

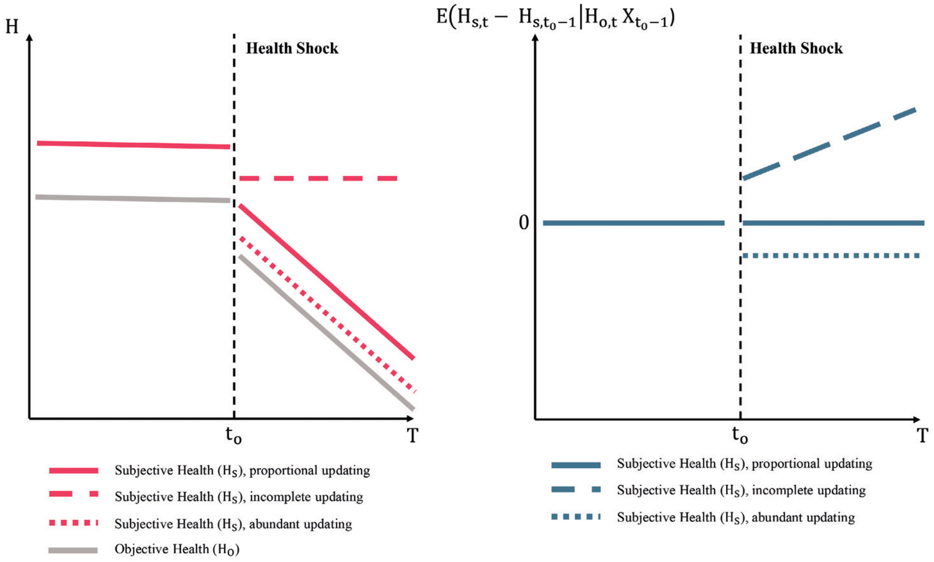
3 Methods

3.1 Conceptual Model and Interpretation

We identify the effect of health events on the perceived health of patients and spouses of patients by exploiting the exogenous timing of severe negative shocks to patients' objective health in the form of heart attacks and strokes. In doing so we aim to identify the magnitude and persistence of learning effects resulting from experiencing health events (patients) or observing them within one's close social network (spouses of patients). Our general approach is motivated by considerations around the different potential mechanisms underlying changes in perceived health around the onset of health events. The most obvious mechanism is the direct impact of a health shock on the objective health of those experiencing the shock. After a severe health shock decreasing objective health (H_o) it is expected that perceived health (H_s) reflects this change. Assuming the general mapping between components of objective health H_o and subjective health H_s to remain persistent before and after the shock one can define an expected perceived health state conditional on objective health $E(H_s | H_o)$ and explore whether realized perceived health changes are proportional to what would be expected given observed objective health after the event. To illustrate our thinking Figure 1 depicts three hypothetical scenarios on the relationship between objective and perceived health around the occurrence of a health shock, panel a), and how these scenarios would be manifested given our empirical model, panel b).

Figure 1: Scenarios on the Effect of Health Shocks on Health Perceptions

a) Objective and Subjective Health Levels   b) Regression Coefficients



Source: Own illustration.

Consider the levels of objective health (grey) and perceived health (red) over time for a hypothetical health shock occurring at time t_o and resulting in an initial decrease and then progressive deterioration of objective health. Panel a) presents three (of the many) possible ways in which an individual might update their health perceptions, each with different implications. The first scenario considers an updating proportional to objective health changes (solid red line). In the alternative scenarios individuals update their health perceptions disproportionate to changes in objective health. A first case represents a form of incomplete updating (dash-dotted red line) where subjective health changes neither fully reflect the immediate changes in objective health at the time of the shock nor the subsequent further deterioration in objective health. The second case on the other hand considers an updating of perceived health that at each point is above what would be expected based on objective health changes (dashed red line).

The interpretation of these discrepancies between subjective and objective health over time differs depending on the type of health change causing them. Consider a road accident that is entirely orthogonal to individual-level characteristics and whose consequence is an increasing level of limited physical mobility with no further implications for other health risks or dimensions. This might be categorized as non-informative beyond the change in current objective health (H_o) and hence health perceptions (H_s) would only be updated proportional to objective health changes (proportional updating, solid green line). However, health shocks that have implications for future health trajectories and survivability, or provide information on pre-existing risk factors, such as heart attacks or strokes, might also have an additional information channel leading to an updating of perceived health beyond what would be expected purely based on objective health changes.

We aim to isolate these different discrepancies between subjective and objective health over time. To do this we use an event-study design in which the last year before the health event (t_0 - 1) is the baseline and perceived health ($H_{s,t}$) is a function of objective health ($H_{o,t}$) and a set of pre-shock demographic and health-related characteristics (X_{t_0-1}). This can be conceptually represented as:

$$(1)\ H_{s,t} = \alpha + \sum_{e=\tau,\ e\neq-1}^{T} \delta_e\, I(t = e) + \rho H_{o,t} + \beta X_{t_0-1} + \varepsilon$$

Where t = τ is the furthest observable relative time-point before the health event and the last relative time-point after the event is t = T. This means that our event-study coefficients δ_e represent the discrepancy between perceived health conditional on two dimensions of objective health information. Current objective health ($H_{o,t}$), and individuals' health and demographic characteristics in the year before the health event (X_{t_0-1}). Accounting for both is important as we are interested whether perceived health changes disproportionate to what would be *expected* given both current objective health and individual-level health information at the time of the event.

Panel b) depicts the coefficient patterns one would observe for each of the three hypothetical scenarios. If updated health perceptions were entirely in line with changes in objective health and pre-existing health information one would expect our event-study coefficient δ_e to be zero (solid blue line) as there are no discrepancies. On the contrary in the case of incomplete updating that both does not reflect changes in objective health at the time of the shock nor the subsequent health trajectory change one would observe a positive and increasing discrepancy as perceived health is *higher* than what would be expected (dash-dotted blue line). Or phrased differently, individuals would have (upward) biased or overly optimistic health perceptions (Arni et al., 2021), as perceived health diverges from objective health. In the alternative case of a decrease in perceived health beyond what would be expected one would observe a downward shift (dashed blue line). This would indicate that perceptions reflect information beyond the pre-existing health information at the time of the shock and the objective health changes that occurred. In our conceptual framework this would indicate learning effects, meaning that new information has been incorporated after the health shock. Such learning effects might be retained persistently or be short-lived, meaning that over time the effect attenuates which could be interpreted as a form of adaptation (Baji & Biro, 2018).

The rationale and conceptual interpretations outlined above for patients experiencing a health shock also apply to spouses of patients (or any other individual not directly impacted by the health event). If there is no direct impact on the physical health of the spouses, discrepancies would only be expected in the form of learning effects. A downward shift in perceived health conditional on spouses' information on their own health before the event and their objective health would be supportive of the hypothesis that health shocks create learning effects (see e.g., Fadlon & Nielsen (2019) or Hodor (2021)). While the absence of such effects in combination with observed behaviour changes with respect to spouses own risky or preventive health behaviours would be supportive of the hypothesis that behavioural spillovers are driven primarily by salience and without direct changes to spouses on perceived health.

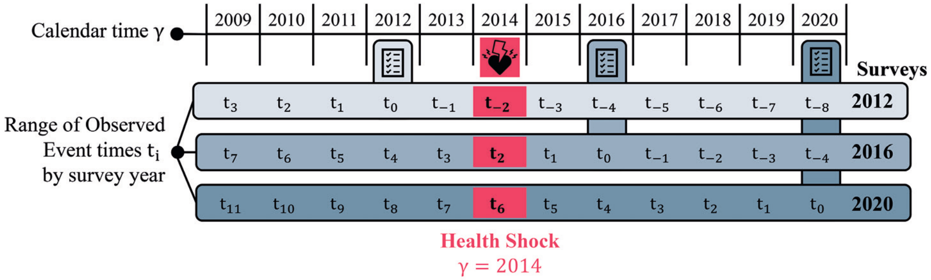
3.2 Study design: Time structure and sample selection

Ideally, we would like to have longitudinal data on measures of objective and perceived health for the same individuals over time, before and after a health shock occurs. Unfortunately, a longitudinal panel survey of sufficient sample size that allows for the identification of heart attack and stroke patients while also following them for long enough time periods to explore dynamic effects for multiple years is not available. Instead, we identify the effect of a heart attack or stroke by comparing the objective and perceived health of individuals in the years right after the health shock to the health of *different* but similar individuals in the years prior to the same shock. The main identifying assumption is that within our study period the timing of the shock is exogenous, so that individuals who have a heart attack or stroke at different times or whose partners have such an event can be compared. This assumption is less stringent than the one needed for a comparison between individuals who receive a shock to individuals that never receive the shock. In the latter case we would have to assume that individuals have no private information about their risk of a heart attack, while in our case people can have this information if they do not know *when* they will have a heart attack.

We use the three health monitor surveys from 2012, 2016, and 2020 linked to administrative records on hospital admissions and their cause at the individual-level for 1995 to 2020 to construct our analysis dataset. By linking hospital records going back to 1995 to the health monitor participants we can identify those individuals that experienced an ischemic heart attack or stroke and completed the health survey at relative time points t_i to their health event. As a result of this we can construct a pseudo-panel based on a repeated cross-section of individuals that are surveyed at different relative times either before the first time they experienced a heart attack or stroke ($t_{<0}$) or after their respective health event ($t_{\geq 0}$). As we later make use of health-related information before the health event for up to three years and our data on individual-level prescription medication use goes back to only 2006 we only consider individuals experiencing their first heart attack or stroke between 2009 and 2020. Online Supplement O4.A1 contains a more detailed description of our health event definition with respect to the specific disease classification codes used to identify the relevant patient samples and how we ensure first-time exposure to a relevant health event.

Figure 2 illustrates how the year of survey participation $t^s \in \{2012, 2016, 2020\}$ and the calendar year of the health event (γ) interact to determine the observed event-time t_i at the individual level. Marked in red is the example case for a health event at γ = **2014** and the resulting observed event-times by survey years. In theory we could construct a repeated cross-section ranging from t_{-8} ($\gamma = 2020$ and $t^s = 2012$) to t_{11} ($\gamma = 2009$ and $t^s = 2020$), we do however collapse the tail-ends of relative event-times resulting in a range of $t_{\leq -4}$ to $t_{\geq 7}$ to account for low cell-populations in these specific observed event-time-groups that are often only sourcing observations from one of the three survey cohorts.

Figure 2: Time Structure by Survey-Year



Source: Own illustration.

While the exogeneity in the timing of health shocks is the main source of exogenous variation we take additional steps to ensure the comparability of the different individuals observed at different event-times. We do so by selecting a sub-sample of individuals based on observable criteria from the universe of first-time admissions. The most salient concern is the role of survival. Individuals interviewed *after* their health event ($t_{\geq 0}$) have survived at least until the time of the survey, individuals interviewed *before* ($t_{<0}$) on the other hand could experience fatal events. We therefore impose a minimum survival period of at least one year to account for this selection effect. We chose this pragmatic basic survival corridor as due to our time-structure we cannot ascertain uniform survival periods without considerable losses of observations due to survival-corridors being truncated (we do however consider the impact of this decision as part of our robustness checks).[9] However, to still account for survival dynamics to some degree we impose an age-at-onset restriction. As longer-term survival is heavily dependent on the age at which a health event occurs, we consider patients in a 5-year age-band around the mean age-at-onset across all patients between 2009 and 2020 to retain a representative sample for the underlying patient population. For heart attacks this is ages 60 to 70 (mean: 65.491) and for strokes 65 to 75 (mean: 70.443).

We also consider only certain households to construct our sample of spouses of patients. Based on municipal records on registered individuals and their relationship to each other we identify those living with a first-time patient at the time of the patients' admission. We only exclude households if there is more than one individual among the group of household-members from our pool of unique patients selected based on the criteria outlined above. This means that all

---

[9] To illustrate, consider two individuals participating in the health monitor of 2020. One suffers from a heart attack in 2013 (observed event-time: t_7), the other in 2019 (observed event-time: t_1). Even if our vital registry data would record deaths occurring after March 2022, which is the last observed time-of-death in the version of this dataset we use, we could only impose a survival period of around four years depending on the exact timing of the health event.

considered spouses do not experience a heart attack or stroke themselves at any time. Lastly, we impose a similar age-at-onset restriction of 5-years around the mean age at onset among spouses of heart attack and stroke patients. This leaves a sample of 55,614 first-time heart attack patients and 44,383 stroke patients that can be followed longitudinally in the administrative data and matched to the pool of health monitor respondents. For spouses these numbers are 42,808 for heart attacks, and 28,908 for strokes.

3.3 Empirical Application

The overarching goal of our empirical strategy is to ensure that despite patients and spouses of patients being exposed to the health event at different time points (admission years) and provide individual survey responses at different relative event-times (t_i) they are comparable at the time of their respective individual shock (t_0). Our selection on age and baseline survival after the event will achieve this in some important dimensions. However, we go further trying to also ensure that individuals compared with each other share a comparable *ex-ante* probability to suffer from the health event or be exposed to it. We do this using the two-step approach that follows the "doubly robust" estimation approach proposed by Callaway and Sant'Anna (2020) and combines inverse-probability weighting with an event-study design (Sant'Anna & Zhao, 2020). This doubly robust procedure combines propensity score-based methods to deal with covariates in event-study regressions proposed by Abadie (2005) with the outcome regression approach (Heckman et al., 1997). The doubly robust property from the fact that correct inference requires only one of the two specifications, the generalized propensity score estimation or the outcome regression, to be correctly specified (Sant'Anna & Zhao, 2020).

We implement a doubly robust design in a two-step procedure. The first step relies on the available longitudinal administrative panel data to estimate the individual-level *ex-ante* probability of experiencing the health event at the eventual individual event-time. A detailed description of this step can be found in Online Supplement O4.A2 alongside information on the balancing of covariates after the matching. Online Supplement O4.A2 also contains a discussion of the underlying identifying assumption outlined by Callaway and Sant'Anna (2020) and their applicability in our context. Broken down to the underlying basic steps we use data on socio-economic (age, gender, labor market status, income, and household characteristics) and rich data on health-related information (medication use by medical conditions and the number of hospitalizations, inpatient days, and types of admissions by disease groups) from the two years prior to the eventual health event to estimate the *ex-ante* probability of experiencing a health shock. To do so we run a logit regression predicting the event-onset in a specific year among a sample of individuals that are affected in that year against all individuals not yet affected by their first-time health event. This is done separately for each health event type (heart attack/stroke) and for the different samples of patients and spouses of patients. The only difference is that for spouses of patients we also include information on their partner's cardiovascular health based on medication use before the event.

From the estimations outlined above we obtain coefficient estimates which are then used to predict a propensity score based on the observed socio-economic and health-related characteristics at the time of the health event for those patients and spouses of patients participating in the health monitor surveys. To ensure that all individuals in our post event-time groups ($t_{\geq 0}$) are surveyed after the health event we exclude all individuals whose health event, or whose partner's health event occurred between September and December, the surveying period of the health monitor, if they are members of the t_0 event time group. This is done as for these individuals there is uncertainty whether they are in surveyed after the event. After imposing an area of common support among the distributions of estimated probabilities we are left with a sample of 7,167 heart attack patients and 5,270 spouses of such patients and 5,836 stroke patients and 3,886 spouses of such patients. See Online Supplement O4.A2 and Table O4.A2.3 for the number of respondents by health event, sample group and survey-year cohort. Importantly, we do not impose that both patients and their spouses answer the health monitor at the same time as the survey is not designed as a household survey. In addition, these numbers reflect the sample of individuals for which we observe both objective and subjective health measures alongside information on pre-treatment health from administrative data. When we later consider specific self-reported health behaviours these numbers are sometimes lower as not all individuals also respond to all questions on self-reported health behaviours.

In a second step we then estimate an event-study regression that uses the estimated propensity scores as weights. We also include control variables in these regressions (hence the term doubly robust) to account for remaining imbalances on key variables. Using this pooled survey data, we regress self-perceived health $H_{s,t}$ (and later health behaviours) on objective health $H_{o,t}$, a set of dummies measuring the time-distance *e* to the health shock $I(t = e)$ at the time of the survey together with a set of control variables:

$$\text{(2) } H_{s,t} = \hat{\alpha} + \sum_{e\leq -4,\ e\neq -1}^{7} \hat{\delta}_e\, I(t=e) + \hat{\rho}H_{o,t} + \hat{\beta}^D X^D_{t_o} + \hat{\beta}^H X^H_{t_o-1} + \hat{\beta}\mathrm{InpatientDays}_{t_o} + \varepsilon$$

Each observation is weighed by the inverse probability estimated using Equation (2). The included control variables cover basic demographic information such as gender and the age in the year at the time of the onset alongside survey-year dummies ($X^D_{t_o}$) and the use of specific medications indicating pre-existing health conditions relevant for cardio-vascular risk in the year before the health event ($X^H_{t_o-1}$). The included medication groups are anti-thrombotic, anti-hypertensive, statins, diabetes-related, treatments for COPD, and medications for acid-related disorders. For spouses of patients, we include information on their own medication use and their partner's medication use in these domains. We also account for potential differences in the intensity of the health event, or the partner's health event, by including information on the number of inpatient days associated with the first-time admission ($\mathrm{InpatientDays}_{t_o}$). We do not include the full range of control variables used for the generalized propensity score model given the comparatively low sample size of the survey data versus the administrative data. Instead, we opt for a parsimonious specification allowing us to address potential remaining differences in the most relevant dimensions of health-information and following the conceptual rationale outline earlier and represented in Equation (1) to capture both information on current latent objective health at the time of the survey response alongside health information individuals have at the time of the health event.

4 Results

4.1 Descriptive Statistics

Table 1 provides an overview of the samples of patients and spouses of patients. For each of these groups it provides summary statistics on demographic and health related characteristics in the year directly preceding the health event. Within each panel for patients or spouses of patients, descriptive statistics are given for the heart attack and stroke samples. For each health event two columns compare the sample of unique health monitor participants against the underlying population of all patients, or spouses of patients, that experience a first-time heart attack or stroke between 2009 and 2020. The health monitor samples are highly comparable to their underlying populations especially with respect to demographic characteristics but most importantly health outcomes before the respective health event. While there are some differences indicating our sample of patients answering the health monitor surveys to consistent to more pensioners overall the survey participants seem to represent a sample representative of the underlying populations of patients and spouses of patients.

4.2 The Impact of Health Shocks on Patients' Health and Health Perceptions

Before we study the development of subjective health *conditional* on physical limitations, we first show the development of physical limitations themselves and of *unconditional* subjective health, in a similar way as in the example in Figure 1 panel a). Figure 3 shows the development of both health measures before and after the health shock (relative to the last period before the shock) for heart attack (a) and stroke (b) patients in our sample. The estimates are derived using the doubly robust estimation method as in Equation (2) but using both perceived health and the absence of any physical functional limitation as the outcome variable and without including dummies for the seven dimensions of functional limitations as a control variable. The vertical axis on the left-hand side depicts the coefficient size for the regression using self-assessed health as the outcome measure. The right-hand side is the coefficient for not reporting any physical limitation. Dots (heart attacks) and diamonds (strokes) indicate the point estimates for the respective specification with bars indicating the 90, 95 and 99% confidence intervals. Effect estimates for perceived health are depicted in color against the estimated effects on the absence of a physical functional limitation in grey. The detailed estimates can be found in Online Supplement Table O4.A3.1.1 (perceived health) and Table O4.A3.1.2 (functional limitations).

Table 1: Descriptive Statistics

	Patients				Spouses of Patients			
	Heart Attack		Stroke		Heart Attack		Stroke	
Demographics	**Survey**	**All**	**Survey**	**All**	**Survey**	**All**	**Survey**	**All**
Age at Onset	65.491	65.033	70.443	70.252	64.530	64.079	67.637	67.375
	(3.093)	(3.145)	(3.048)	(3.120)	(3.081)	(3.134)	(3.048)	(3.117)
Female	0.253	0.270	0.402	0.416	0.740	0.757	0.620	0.644
	(0.435)	(0.444)	(0.490)	(0.493)	(0.439)	(0.429)	(0.485)	(0.479)
Living Alone	0.180	0.205	0.234	0.280	-	-	-	-
	(0.383)	(0.403)	(0.423)	(0.449)	-	-	-	-
Pensioner	0.612	0.535	0.935	0.919	0.428	0.337	0.672	0.579
	(0.487)	(0.499)	(0.245)	(0.272)	(0.495)	(0.472)	(0.470)	(0.494)
Incapacity Benefit	0.058	0.082	0.007	0.014	0.053	0.069	0.036	0.051
	(0.235)	(0.275)	(0.085)	(0.118)	(0.225)	(0.252)	(0.187)	(0.219)
Per-capita Household Income	2.741	2.639	2.300	2.197	2.713	2.661	2.367	2.300
(monthly, 1000€)	(2.218)	(2.064)	(1.907)	(1.981)	(2.211)	(2.039)	(1.857)	(1.879)
Health in the year before Admission								
Hospitalisations	0.288	0.307	0.366	0.381	0.248	0.253	0.280	0.292
	(1.053)	(1.371)	(1.062)	(1.148)	(1.023)	(1.023)	(0.927)	(1.178)
Total Length of Stay	0.783	0.879	1.065	1.241	0.563	0.656	0.771	0.816
	(3.676)	(4.280)	(4.352)	(5.080)	(2.300)	(3.496)	(3.824)	(3.996)
Medication: Anti-Thrombotic	0.246	0.254	0.328	0.334	0.163	0.154	0.217	0.214
	(0.431)	(0.435)	(0.469)	(0.472)	(0.369)	(0.361)	(0.412)	(0.410)
Medication: Anti-Hypertensive	0.442	0.438	0.516	0.533	0.356	0.340	0.413	0.405
	(0.497)	(0.496)	(0.499)	(0.499)	(0.479)	(0.474)	(0.493)	(0.491)
Medication: Statins	0.299	0.300	0.354	0.353	0.237	0.219	0.291	0.281
	(0.458)	(0.458)	(0.478)	(0.478)	(0.425)	(0.413)	(0.454)	(0.450)
Medication: Diabetes	0.115	0.144	0.159	0.182	0.082	0.085	0.101	0.109
	(0.312)	(0.351)	(0.365)	(0.386)	(0.274)	(0.279)	(0.302)	(0.312)
Medication: COPD	0.129	0.142	0.158	0.158	0.119	0.129	0.126	0.135
	(0.336)	(0.349)	(0.365)	(0.365)	(0.324)	(0.335)	(0.332)	(0.342)
Medication: Acid-related disorders	0.266	0.287	0.321	0.325	0.243	0.236	0.268	0.263
	(0.442)	(0.452)	(0.467)	(0.468)	(0.429)	(0.425)	(0.443)	(0.440)
Observations	**7,161**	**55,614**	**5,831**	**44,383**	**5,625**	**42,808**	**3,883**	**28,908**

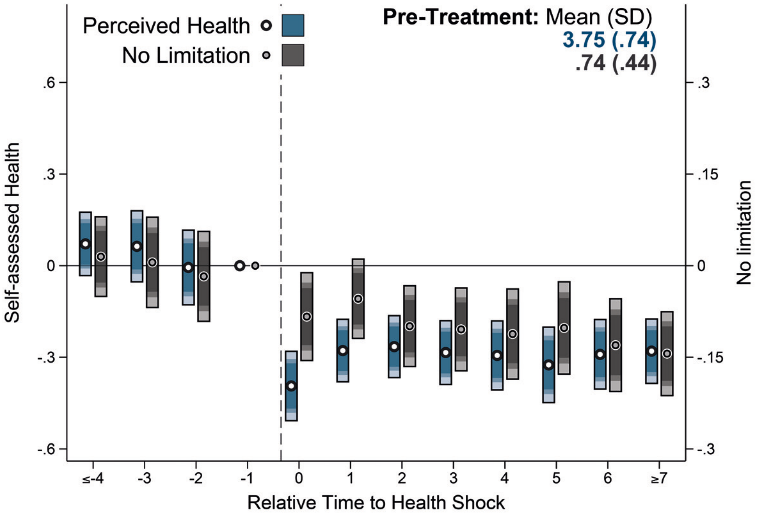
Source: Own calculations based on data from Statistics Netherlands. *Note:* Depicted are the sample characteristics for unique patients and spouses of patients that experience a heart attack/stroke for the first time between 2009 and 2020 at the time of the health event occurred. It contrasts the characteristics among the respective sample of Health monitor survey participants, labelled “survey”, against the underlying full population sample of unique patients or spouses of patients, labelled “all”.

For both heart attacks and strokes, we do not see a significant trend in objective or subjective health prior to the shock but a clear effect on both with respect to post-shock trajectories. Both heart attacks and strokes result in an initial decrease in the share of individuals reporting to not suffer from a physical limitation that gradually decreases over time. However, strokes result in an overall larger decrease in objective health while also affecting different dimensions. Heart attack patients are 8.34% less likely to report no level of physical limitation immediately after the admission at t_0, rising to 14.4% for the group of patients interviewed at $t_{\geq 7}$ after and more. For strokes these numbers are initially smaller at 7.61%, rising to 27.2%. This general pattern of strokes resulting in more severe objective health decreases over time also holds when considering a higher threshold for a functional limitation to be reported (see Online Supplement Figure O4.A3.1.1 and Table O4.A3.1.3) while they also have a more pronounced effect on cognitive limitations, especially speaking and hearing related limitations (see Online Supplement Figures O4.A3.1.2 and O4.A3.1.3). We also consider the broader effects on long-term risk of re-occurring cardiovascular health events and related or general inpatient hospital admissions. Online Supplement Table O4.A3.1.4 provides the detailed results and in line with the effects for functional limitations we find also long-term increases in hospital care use, indicating substantive and persistent health effects both on latent health and general health risk.

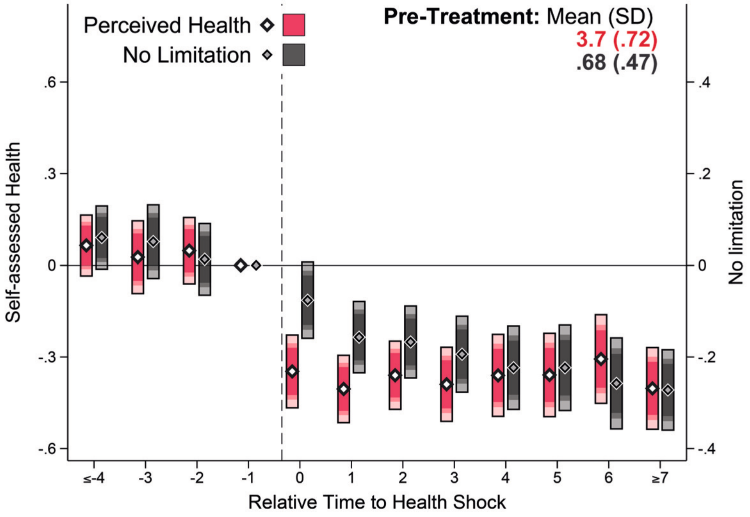
With respect to subjective health, we find strong and persistent effects on perceived health. For heart attack patients perceived health is lower by 0.394 points immediately after the event (t_0), decreasing slightly to 0.278 in the subsequent year (t_1) and remaining stable thereafter. For strokes on the other hand the initial effects are slightly smaller at -0.347 (t_0) remaining roughly stable ranging between -0.405 (t_1) and -0.307 (t_6). In summary we therefore observe strong effects of both health events on patient's objective and subjective health across dimensions. However, while the effects on objective health seem to be increasing over time the effects on subjective health take a different pattern, suggesting no evidence for a reversion or adaptation in perceived health over time and a persistent level shift in perceived health.

Figure 3: The Effect of Health Shocks on Perceived Health and the Absence of any Physical Functional Limitation

a) Heart Attack



b) Stroke



Source: Own calculations based on data from Statistics Netherlands. *Note:* Event study estimates of the effect of a heart attack/stroke on self-perceived health and the absence of a reported physical limitation. Dots/diamonds depict point-estimates with shaded bars indicating the 90/95/99 confidence intervals. Self-perceived health is a five-point measure ranging from 1 *"Very Bad"* to 5 *"Very Good"*. Absence of a physical limitation is defined as reporting to be in no way limited to doing any of the following activities: carrying 5kg for 10 meters, being able to reach for the ground, walking for 400 meters without stopping. Included controls are age-at-onset (2-year dummies), sex, income-quintile membership the year before admission, pension/incapacity benefit receipt the year before admission, medication use the year before admission (anti-thrombotic, anti-hypertensive, statins, diabetes, COPD, acid-related disorders), inpatient days associated with the heart attack/stroke admission, and survey-year dummies.

The Effect of Health Shocks on the Relationship between Perceived and Objective Health

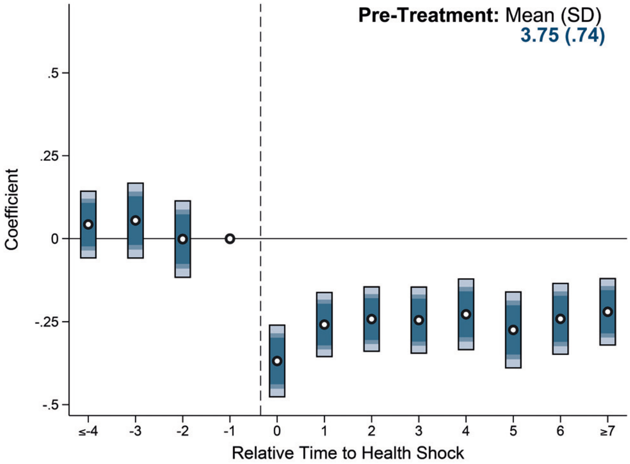
We next turn to exploring the relation between objective and subjective health after the health shock, by turning to the event-study of subjective health conditional on objective health, as in Equation (2). Figure 4 depicts the estimated event-study coefficients $\hat{\delta}_e$ separately for the heart attack (a) and stroke (b) samples.

Comparing the results of the effect of a health shock on perceived health without conditioning on the different dimensions of functional limitations (Figure 3) to the results after conditioning on objective health (Figure 4) we generally observe the same pattern. Health shocks result in a substantial and persistent adjustment in perceived health, suggesting that there is no adaptation and that instead health events result in a long-term downward adjustment to subjective health beyond changes in objective health. While conditioning on objective health does not alter this pattern it does however generally decrease the observed effect sizes although not substantially. For heart attacks the coefficients decrease from -0.394 without conditioning on objective health to -0.368 (t_0) and for later years more substantially e.g., at $t_{\geq 7}$ from -0.280 to -0.220. For strokes the overall effect sizes differences between the specifications are larger than for heart attacks but once we account for objective health for both events the effect estimates for perceived health changes are highly similar both in their pattern and general magnitude.

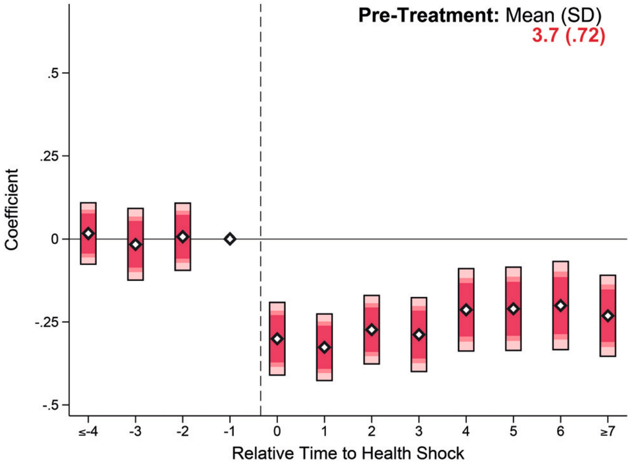
As we measure perceived health using self-assessed health on a categorical scale from 1 *"Very Bad"* to 5 *"Very Good"* we also conduct an alternative analysis using an ordered logit model. This allows us to translate the linear coefficient estimates into more informative marginal changes in the response probabilities for specific response categories. Online Supplement Figure O4.A3.1.4 provides a visualisation of these results. For both health shocks we find the effects on perceived health to be substantial. As an illustrative example, among heart attack patients, the cumulative share of the two highest response categories was 65.62 percent, split between 55.61 percent indicating *"Good"* and 11.01 percent indicating *"Very Good"* health in the year before their heart attack. The estimated change in the marginal response categories implies substantial effects at t_0 with a decrease by 16.16 percentage points (or 29.11%) for the *"Good"* and 6.49 percentage points (or 58.94%) for the *"Very Good"* response category. While these initial effects decrease in subsequent years to around 10 percentage points and 4 percentage points respectively, they remain substantial, and the general patterns are similar for stroke patients.

Figure 4: The Effect of Health Shocks on Perceived Health conditional on Objective Health

a) Heart Attack



b) Stroke



Source: Own calculations based on data from Statistics Netherlands. *Note:* Event study estimates of the effect of a heart attack/stroke on self-perceived health. Dots/diamonds depict point-estimates with shaded bars indicating the 90/95/99 confidence intervals. Self-perceived health is a five-point measure ranging from 1 *"Very Bad"* to 5 *"Very Good"*. Included controls are age-at-onset (2-year dummies), sex, income-quintile membership the year before admission, pension/incapacity benefit receipt the year before admission, medication use the year before admission (anti-thrombotic, anti-hypertensive, statins, diabetes, COPD, acid-related disorders), inpatient days associated with the heart attack/stroke admission, survey-year dummies, and dummies for each of the seven functional limitation dimensions being reported to be heavily or entirely limited.

4.3 The Impact of Health Shocks on Health Behaviours

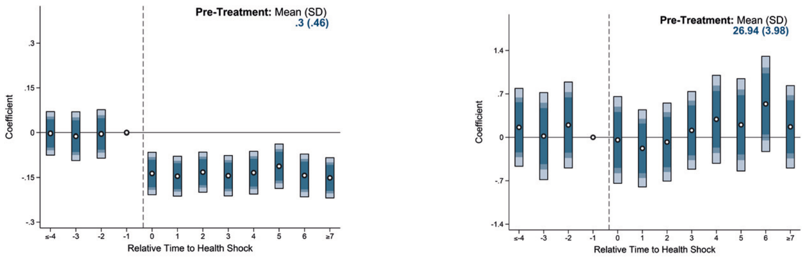
Our results indicate that health shocks result in substantial and long-term persistent changes in perceived health following a health shock and that these changes in perceived health go beyond the experienced changes in objective health. Such a considerable adjustment even after accounting for objective health changes suggests patients to update their health perceptions, incorporating new information in relation to the experienced health shock, and in line with this one would also expect adjustments in health behaviours. To explore this, we will now turn towards the effects of health shocks on health behaviours, first considering self-reported health behaviours and then adherence to preventive medication regiments.

Self-Reported Health Behaviours

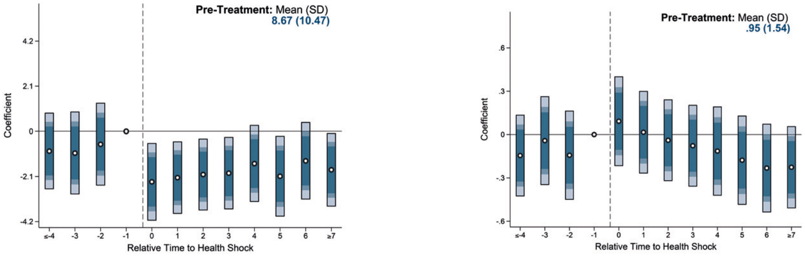
Figure 6 provides an overview of our results with respect to the effect of health shocks on four different health behaviours: smoking, alcohol consumption, bodyweight, and exercising habits. Our results suggest significant behavioural adjustments to occur after heart attacks and strokes, but these are limited to the most high-risk health behaviours in the context of cardio-vascular health, smoking and alcohol consumption. Experiencing either a heart attack or stroke results in large and persistent changes in smoking prevalence. Heart attacks result in an immediate decrease in the smoking rate of 13.66 percentage points (a 45.53% decrease compared to the pre-treatment mean) at t_0 remaining persistent and even slightly increasing for the cohort of individuals interviewed seven years or more after their health event. For strokes the immediate effects are smaller with 5.22 percentage points decrease and only significant at the 90% level. However, subsequent effects increase gradually indicating a persistent 10 percentage point decrease in the smoking prevalence (a 40% decrease compared to the pre-treatment mean). With respect to alcohol consumption, we observe a similar pattern although overall the results are of lower magnitude compared to the impact on smoking rates. Heart attacks result in an immediate decrease in the number of weekly drinks of 2.35 (a 27.33% decrease compared to the pre-treatment mean) and strokes in a decrease of -2.16 drinks (or a 22.93% decrease compared to the pre-treatment mean). For both these effects are generally persistent with no clear pattern indicating a reversion in the behavioural adjustments. Contrary to the large and persistent effects on smoking prevalence and alcohol consumption we observe no effects on exercising habits or body-mass-index for either the heart attack or stroke samples.

Figure 5: The Effect of a Health Shock on Self-reported Health Behaviours

a) Heart Attack

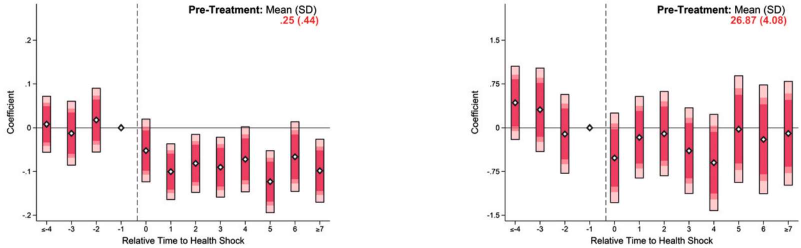

Currently smoking

Body-Mass-Index

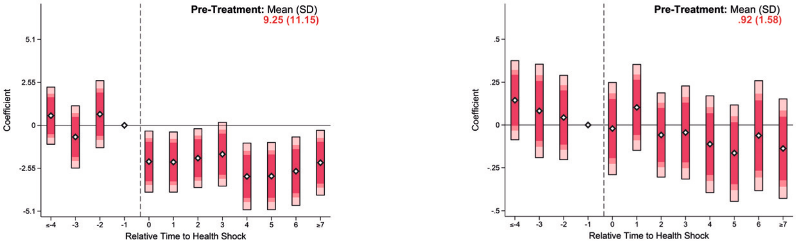

Number of alcoholic drinks a week

Number of days a week exercising

b) Stroke


Currently smoking

Body-Mass-Index


Number of alcoholic drinks a week

Number of days a week exercising

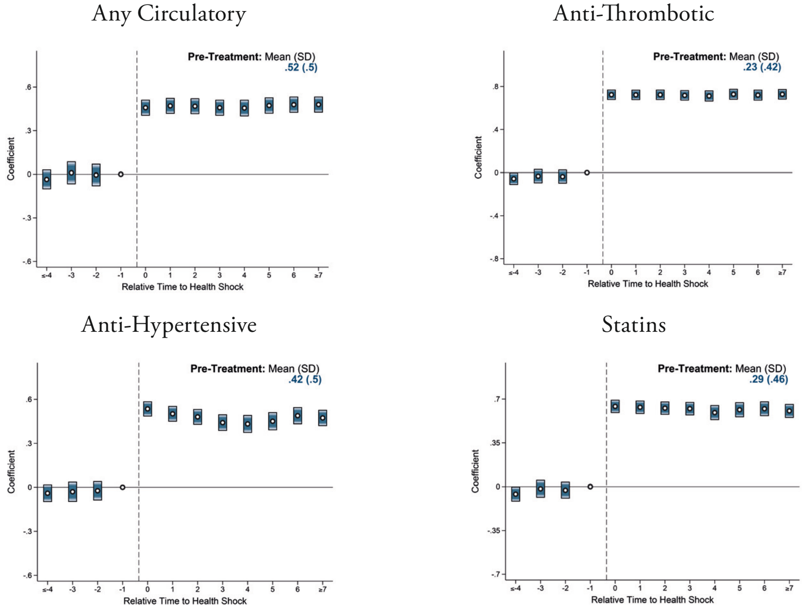
Source: Own calculations based on data from Statistics Netherlands. *Note:* Event study estimates of the effect of a heart attack/stroke on self-reported health behaviours. Dots/diamonds depict point-estimates with shaded bars indicating the 90/95/99 confidence intervals. Included controls are age-at-onset (2-year dummies), sex, income-quintile membership the year before admission, pension/incapacity benefit receipt the year before admission, medication use the year before admission (anti-thrombotic, anti-hypertensive, statins, diabetes, COPD, acid-related disorders), inpatient days associated with the heart attack/stroke admission, survey-year dummies, and dummies for each of the seven functional limitation dimensions being reported to be heavily or entirely limited.

Preventive Prescription Medication Use

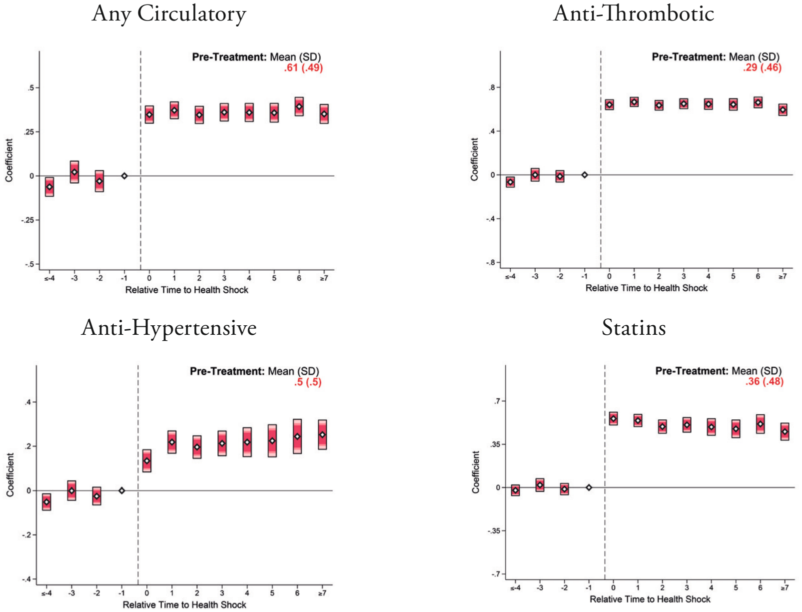
After considering self-reported health behaviours we now consider the use of cardio-vascular specific preventive medications: anti-thrombotic, anti-hypertensive and statins. Figure 7 presents our effect estimates of the effect of a heart attack or stroke on taking either of these types of medication and then separately for each medication type. Overall, our results provide a clear and consistent picture. While at the baseline pre-treatment period already about 51.60% of heart attack and 60.30% of stroke patients take either of the three preventive medication types this increases by 45.80 percentage points (heart attack) and 34.80 percentage points (strokes) for the year immediately after the health event (t_0). This is a near universal uptake of at least one type of preventive medication with effects remaining stable throughout, indicating long-term adherence. When considering the different types of medication this pattern is mirrored. Although there are some differences in the relative effect sizes these are mainly driven by differences in the baseline consumption levels. Experiencing a heart attack or stroke results in an initial uptake of anti-thrombotic medications and statins resulting in roughly 90% or more of patients consuming these medications and adhering to a long-term medication regiment. Only for anti-hypertensive medication we see a smaller effect for strokes, but not for heart attacks, bringing the consumption levels to only 75%. For strokes we also observe a slight reversion of the initial effects, but this is limited to statin use where our estimated effects suggest an initial increase by 55.80 percentage points (t_0), falling gradually to 45.20% ($t_{\geq 7}$).

Figure 6: The Effect of a Health Shock on Prescription Medication Use

a) Heart Attack



b) Stroke



Source: Own calculations based on data from Statistics Netherlands. *Note:* Event study estimates the effect of a heart attack/stroke on prescription medication use. Dots/diamonds depict point-estimates with shaded bars indicating the 90/95/99 confidence intervals. Included controls are age-at-onset (2-year dummies), sex, income-quintile membership the year before admission, pension/incapacity benefit receipt the year before admission, medication use the year before admission (anti-thrombotic, anti-hypertensive, statins, diabetes, COPD, acid-related disorders), inpatient days associated with the heart attack/stroke admission, survey-year dummies, and dummies for each of the seven functional limitation dimensions being reported to be heavily or entirely limited.

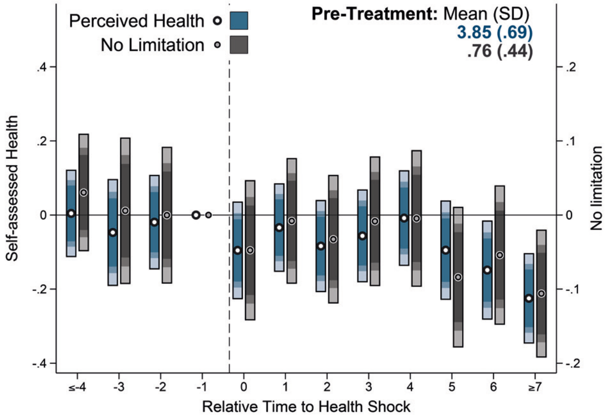
4.4 Spillovers of Health Shocks on Spouses

Perceived Health and Health Trajectories of Spouses

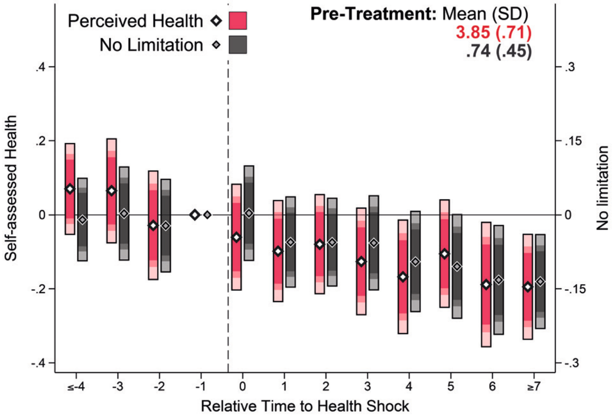
In the second part of our analyses, we now turn to examine spillover effects of health shocks on the health perceptions and health behaviours of spouses of patients. We first replicate the analyses underlying Figure 3 and Figure 4 for our sample of spouses of patients (the corresponding results for perceived health are found in Online Supplement Table O4.A3.2.1). Figure 7 presents our estimates for the effect of the partner's health event on the perceived health (colored) and the reported absence of a physical functional limitation (grey). In contrast to our results for patients themselves we find no evidence that a partner's severe health event results in changes to the spouse's perceived health or their objective health. While all event-time coefficients after the health shock ($t_{\geq 0}$) are negative they are not significant beyond the 90% level. Also, for the absence of a physical limitation there is no evidence for a short- or medium-term effect of the partner's health shock on the spouse's health and neither do we find such effects on other functional limitation dimensions (for details see Online Supplement Tables O4.A3.2.2 and O4.A3.2.3). These findings suggest that at least for the immediate time after a patient's health shock spousal health perceptions and objective health are not impacted. However, looking at longer-term effects there do seems to be emerging a significant difference in both perceived health and objective health that gradually emerges over time, especially with respect to spouses of stroke patients and to a lesser degree for spouses of heart attack patients.

Figure 7: The Effect of Health Shocks on Perceived Health and the Absence of any Physical Functional Limitation

a) Heart Attack



b) Stroke



Source: Own calculations based on data from Statistics Netherlands. *Note:* Event study estimates of the effect of a heart attack/stroke on spouses' self-perceived health and the absence of a reported physical limitation. Dots/diamonds depict point-estimates with shaded bars indicating the 90/95/99 confidence intervals. Self-perceived health is a five-point measure ranging from 1 *"Very Bad"* to 5 *"Very Good"*. Absence of a physical limitation is defined as reporting to be in no way limited to doing any of the following activities: carrying 5kg for 10 meters, being able to reach for the ground, walking for 400 meters without stopping. Included controls are age-at-onset (2-year dummies), sex, income-quintile membership the year before admission, pension/incapacity benefit receipt the year before admission, medication use the year before admission (anti-thrombotic, anti-hypertensive, statins, diabetes, COPD, acid-related disorders), inpatient days associated with the heart attack/stroke admission, and survey-year dummies.

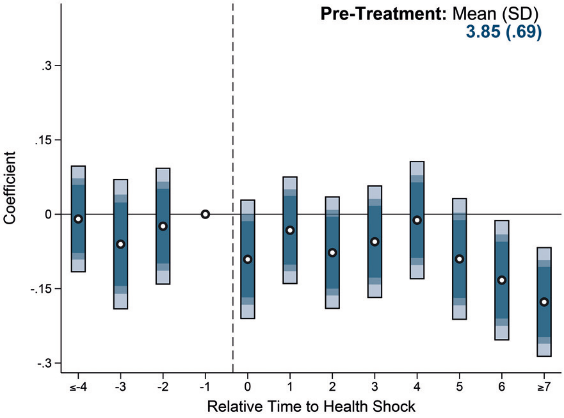
Figure 8 presents our estimates of the effects of a partner's health shock on the spouse's health perceptions when conditioning on objective health at the time of the survey, therefore estimating Equation (2) for spouses. As in the case of patients accounting for objective health at the time of the survey generally lowers the estimated coefficients. In the case of stroke patients, it also attenuates the apparent trend in the post-shock event-study coefficients that would indicate the effect on spouses' perceived health to increase over time. Generally, however, we find no evidence for a consistent short- or medium-term effect of a partner's health event on the spouse's perceived health after accounting for differences in objective health. The previously observed effect estimates for later event-study coefficients ($t_{\geq 6}$) for heart attacks do however remain significant while for the stroke sample throughout coefficients are negative with some being significant. As for our patient sample we translate these average effects into marginal changes in the response probability across response categories, these results can be found in Figure O4.A3.2.1 in the Online Supplement. In summary, for the period directly after the partner's health event, arguably when its occurrence is the most salient to the spouse, we find no conclusive evidence that their perceived health deviates from what would be expected given their health at the time of the survey and their health at the time of the partner's heart attack or stroke.

Self-Reported Health Behaviours and Medication Use of Spouses

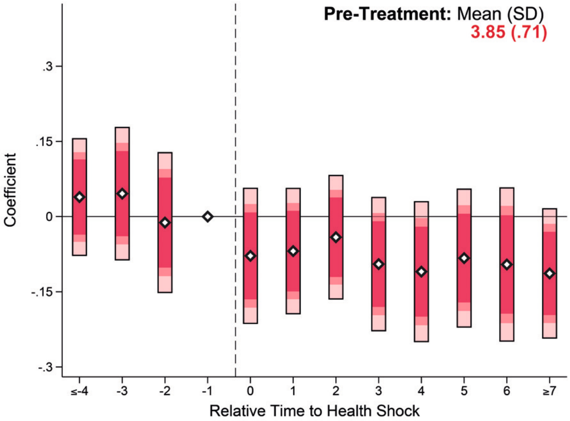
Lastly, we turn towards the spillover effects on spouses' health behaviours and medication use. For brevity we only shortly discuss these results qualitatively. As for patients we first consider the four self-reported measures of risky and preventive health behaviours: current smoking status, alcohol consumption, body-mass-index, and exercising habits. The results from these estimations are found in Online Supplement Table O4.A3.2.5 while Figure O4.A3.2.2 provides the corresponding event-study plots, both can be found in the Online Supplement. For each of these dimensions we find no evidence for a partner's health shock to impact spouses' health behaviours. Solely for spousal smoking behaviour there are some event-time coefficients significant, but only for the stroke sample. We also consider prescription medication use among spouses for either anti-thrombotic, anti-hypertensive, statins, or either of these types of medications as we do for the patient sample. Online Supplement Table O4.A3.2.6 covers the detailed results from these analyses while the corresponding event-study plots can be found in Figure O4.A3.2.3. For medication use our effect estimates generally suggest a strong effect of the partner's health shock on spouses' medication consumption throughout all four measures. Unlike for patients these estimates do not suggest an immediate effect but rather a trend-shift in medication use with an increasing uptake over time. However, the event-study plots also suggest that these results could be explained by an existing pre-trend. For all medication outcomes we observe a consistent picture of significant differences in among the survey participants interviewed four years before the health event ($t_{\leq -4}$) and an overall increasing trend in the coefficient estimates. We therefore interpret these results with a dose of skepticism and explicitly explore whether pre-trends could explain these results in our robustness checks.

Figure 8: The Effect of Health Shocks on Spouses' Perceived Health conditional on Objective Health

a) Heart Attack



b) Stroke



Source: Own calculations based on data from Statistics Netherlands. *Note:* Event study estimates of the effect of a heart attack/stroke on spouses' self-perceived health. Dots/diamonds depict point-estimates with shaded bars indicating the 90/95/99 confidence intervals. Self-perceived health is a five-point measure ranging from 1 *"Very Bad"* to 5 *"Very Good"*. Included controls are age-at-onset (2-year dummies), sex, income-quintile membership the year before admission, pension/incapacity benefit receipt the year before admission, medication use the year before admission (anti-thrombotic, anti-hypertensive, statins, diabetes, COPD, acid-related disorders), inpatient days associated with the heart attack/stroke admission, survey-year dummies, and dummies for each of the seven functional limitation dimensions being reported to be heavily or entirely limited.

5 Robustness Checks

To explore the robustness of our results to certain identification threats, alternative explanations, or modelling choices we conduct a range of supplementary analyses. Firstly, we consider the role of health events affecting health perceptions through their potential impact on mental health, in particular depression or anxiety. Poor mental health, or states of heightened emotional agitation, have been found to be correlated with individual level decision making through their role in shaping individuals' risk preferences and general expectations (see for example recently Meier (2021) or Chuong et al. (2022)). In the context of this study higher levels of depression and anxiety following the experience of a severe health event or a partner experiencing such an event could also affect perceived health. Phrased differently, lower levels of mental health could manifest themselves by altering the mapping between objective health and perceived health by resulting a more negative evaluation of perceived health despite no change oi. To explore this, we estimate the effects of health shocks on the mental health of patients and spouses of patients directly by considering two different mental health outcomes, the Kessler 10-item psychological distress scale (Kessler et al., 2002), a clinically validate screening tool for depression and anxiety related disorders, and prescription medication use for these conditions. The results of these analyses are presented in Online Supplement O4.A4.1 for patients and spouses of patients. We find no evidence for negative mental health effects as the main mechanism for our results on perceived health.

A second concern is the role of other unobserved confounders that could underly the observed pre-trends for certain specifications e.g., in our results for spouses' prescription medication consumption. A particular concern is here that we cannot include admission-year fixed effects alongside survey-year fixed effects and event-time dummies as we rely on a repeated cross-section and not individual-level longitudinal data. As a result of this the observed pre-trends could stem from unaccounted changes over time, like an increasing usage of prescriptive medications in the healthcare system more broadly. To explore whether the observed pre-trend is sufficient to explain the estimated treatment effect estimates we follow Freyaldhoven et al. (2021) by rescaling the coefficients as deviations from an extrapolated linear pre-trend. Online Supplement Figures O4.A4.2.1 and O4.A4.2.2 depict the baseline estimates and the extrapolated linear pre-trend based on these estimates against a set of adjusted coefficients for an exemplary selection of our main outcome measures. For most of our findings there is no evidence of such a pre-trend driving the results, the only notable exception being spouses' medicine consumption. After adjusting the coefficients, most estimated effects become insignificant.

Thirdly we consider different survival corridors than the minimum survival period of one year imposed so far. The rationale for this is that selective survival of healthier individuals could bias our results towards finding long-term persistent effects, especially for patients. To address this, we expand the survival corridor stepwise from one year to at least five years by excluding all patients and spouses of patients that die within this expanded survival corridor. We further exclude individuals for whom this survival corridor cannot be ascertained given that we can only observe survival up until March 2022 but include individuals whose health event only occurs in 2020. Online Supplement O4.A4.3 provides an overview of the results for both patients and spouses of patients for different survival corridors for an exemplary subset of our main outcome measures. Besides an increase in the general uncertainty of our estimates due to a lower sample size the results remain largely unchanged.

Lastly, we consider the importance of the choice of different propensity scores used in the inverse probability weighting for the obtained estimation results. As described in Online Supplement O4.A2 we estimate separate prediction models for each of the health monitor groups and then predict propensity scores separately before pooling the sample. To explore the sensitivity of our results to the choices of which generalized propensity score model is used, we compare our results if instead we use only one of these models estimated for the 2012, 2016, or 2020 health monitor samples to predict propensity scores for all or conduct no weighting whatsoever, therefore only using an outcomes-regression based approach. Online Supplement O4.A4.4 depicts the results of this exercise, indicating no difference in our overall conclusions.

6 Discussion & Conclusion

Experiencing health events directly or observing them within one's social network can be a potent channel in the formulation of health beliefs and behaviours. This paper investigated the role of sudden and severe health events in shaping health perceptions and behaviours of patients, and the spillover effects of such events on the perceptions and behaviours of spouses of patients. By considering how the experience of a health event, either directly (patients) or indirectly (spouses of patients), affects perceived health conditional on objective health dimensions and health information existing at the time of the event we explore whether such events result in learning effects and related behavioural adjustments. Our identification strategy relies on a combination of cross-sectional survey data from the Netherlands linked to detailed longitudinal administrative data on health and socio-economic information around the onset of health events. This unique combination allows us to construct samples of patients and spouses of patients suffering from a heart attack or stroke that are exposed to similar health events and are highly comparable with respect to their characteristics at the time of the event but are surveyed at different relative time points. We therefore exploit the exogeneity in the timing of heart attacks and strokes using a treated-vs-not-yet-treated event study design for repeated cross-sectional data (Callaway & Sant'Anna, 2020) to obtain causal estimates on the long-term effects of such events on health perceptions and behaviours.

Experiencing a severe health event results in long-term persistent changes to patients' perceived health. For both health events we observe a similar downward adjustment in perceptions that persists nearly unchanged for up to seven years after the event. Importantly, these patterns also remain when incorporating information on patients' objective health at the time of the event alongside information on their current objective health in the form of multiple dimensions of health-related physical and cognitive/sensory functioning. This suggests that adjustments to perceived health go beyond what would be expected given the deterioration of objective health after a heart attack or stroke or the level of pre-existing health information. Instead, these changes indicate long-term adjustments in the form of learning effects. We therefore find limited evidence supportive of previous studies suggesting that health perceptions are only temporarily affected by severe health events (Baji & Biro, 2018). This also implies that the findings on adaptation in perceived health after health changes, in particular the onset of functional limitations (see recently de Hond et al. (2019) or Stöckel et al. (2023)), might not be applicable in contexts where these changes stem from health events carrying distinct informational value on future health outcomes. This suggests that adaptation in health perceptions, if it occurs, is limited in scope, and only observed for certain health conditions, in line with indicative evidence from previous studies (see e.g., Cubí-Mollá et al., 2017).

In line with our results on how changes in perceived health indicate long-term persistent learning effects, we also find evidence for behavioural adjustments consistent with this. Smoking and alcohol consumption remain the most important avoidable risk factors for cardiovascular health (Roth et al., 2020). Our results show that following heart attacks and strokes patients decrease their exposure to avoidable health risks and increase their investment in high-value prevention efforts substantially. Our effect estimates suggest for both patient groups a near halving of the smoking rate with a decrease of around 10-15 percentage points against a baseline smoking rate of 24-30%. For alcohol consumption these effects are smaller, equivalent to a 25% decrease in the number of weekly drinks compared to the baseline, but equally persistent. With respect to exercising habits and bodyweight management we observe no evidence of meaningful changes following a health event. Lastly, we observe a near universal and persistent uptake of preventive medications in the form of anti-thrombotic and anti-hypertensive medications, and statins. While initial uptake effects are somewhat mechanical given clinical guidelines around the usage of such medications to decrease risks of subsequent heart attacks or strokes (Binno, 2016) the persistence of the effects on medication use is in line with the effects observed on perceived health.

While we find a set of clear and consistent results suggesting that patients make economically relevant adjustments to their health behaviours likely driven by persistent changes to their health beliefs our results for the spouses of patients are less clear cut. Taken at face value, our results run contrary to the existing literature finding behavioural spillovers for spouses to be substantial, especially with respect to healthcare use (Fadlon & Nielsen, 2019; Hodor, 2021; Hoagland, 2021). We find neither consistent evidence for decreased engagement in risky health behaviours nor an uptake of preventive care consumption, like prescriptive medication use. While these results contradict previous findings, they are in line with our results on the causal effect of a partner's health event on the spouse's perceived health which indicates no significant learning effects. There are few existing studies estimating the causal effect of spillovers in both health perceptions and self-reported health behaviours that can serve as a comparison. Those that provide descriptive evidence on intra-family spillovers on both perceptions and behaviours suggest such spillovers to be limited in scope (Darden & Gilleskie, 2016). More recent evidence on the case of spousal spillovers from diabetes diagnosis also suggests that spillovers, if they occur, might be driven by joint household decision making as opposed to direct (social) learning effects (Thomas & Mentzakis. 2022).

However, while internally consistent, our results on the spouses of patients should be considered with some skepticism. Firstly, all results are based on the analysis sample of health monitor participants and not based on analysing the individual-level longitudinal changes for the whole population of spouses of patients, or other members of a patient's wider social network, as done by Fadlon & Nielsen (2019). In the robustness checks we illustrate that after accounting for observed pre-trends resulting from this reliance on repeated cross-sectional data, we do not find evidence of medication uptake of spouses. Our reliance on relatively small samples of survey participants could however also result in a lack of statistical power. Fadlon & Nielse (2019) report numerically small but economically relevant changes in spouses' statin consumption given the baseline against which they occur. This could be addressed by not imposing the same sample selection criteria with respect to the age-at-onset on spouses as we do for patients or by expanding our sample to also consider adult children and other social network members. Secondly, we currently only consider medication uptake as the main form of preventive care consumption.

Prevention efforts, however, could also be observed with respect to other types of care, most notably increased health information seeking through frequent GP visits or other types of preventive care consumption. Therefore, the lack of observed spillovers among spouses could at least in part be driven by certain methodological choices and the focus on specific outcomes only.

Nonetheless, our results have a range of implications for future research and policymaking. Firstly, while our results indicate substantial learning effects on average, we do not consider whether these effects differ across subgroups. Heterogenous learning is particularly interesting given that despite substantial and persistent behavioural adjustments, like the smoking prevalence dropping by nearly 50% according to our estimates, a non-negligible share of individuals continues to engage in risky behaviours even after a heart attack or stroke has occurred. The underlying mechanisms explaining this, and their implications, are varied. Learning effects could differ across gender, socio-economic, or education groups or be related to pre-existing differences in the accuracy of health beliefs (Arni et al., 2021), thereby increasing health-related inequalities. Continuing to engage in risky health behaviours could also be driven by substitution between different dimensions of behavioural prevention efforts (Kaestner et al., 2014), for example using prescription to avoid the (short-term) disutility resulting from behavioural adjustments and individual's consumption preferences (Darden, 2017). Understanding these heterogeneities would help in identifying patients whose window-of-opportunity for behaviour change is narrow or even shut, in the temporal dimension and/or limited to specific behaviours. These insights could help informing the design of targeted interventions steering individuals towards high-value care consumption matching the health information they should obtain from experienced health events (Fadlon & Nielsen, 2019; Hoagland, 2021).

A related direction for future research relates to key assumptions underlying our empirical approach and the interpretation of results. We interpret the estimated effects on perceived health as learning effects by assuming that the relationship between dimensions of objective health and pre-existing health information at the time of the event remains unchanged by the event itself. However, a change in this relationship could itself be an indication of learning effects as individuals differently weight dimensions of objective health information in the subjective evaluation of their health. This could for example be explored by estimating cut-point-shifts to explore the importance of reference points and scale recalibration (Groot, 2000; Lindeboom & van Doorslaer, 2004). Further disentangling mechanisms underlying the learning effects would also be interesting with respect to the degree that individuals incorporate beliefs about future health events as opposed to pre-existing information and thereby reveal private information (Finkelstein & McGarry, 2006). Studies suggest that perceived health contains substantial information on future healthcare needs and is often a better predictor of mortality unless sufficient (longitudinal) data on objective health dimensions is available (Nielsen, 2016). The degree to which the estimated learning effects are thereby driven by the re-evaluation of existing information versus learning about future health risks would itself be interesting. These dynamics also have implications with respect to heterogenous learning effects discussed earlier, and their influence economic decisions based on health beliefs and private information, such as optimal health insurance choices (see e.g., Handel et al., 2020).

In summary, this paper has identified the long-term persistent causal effect of health shocks on the health perceptions and behaviours of patients. These findings underline the role played by health events in shaping beliefs and behaviours over the long-term and suggests that how individuals think about their health, and how these perceptions are shaped, plays a crucial role in determining persistent health behaviour changes. At the same time our results on the spillovers of such events on spouses of patients provides only limited evidence for broad behaviour change due to social learning effects. Taken together these results highlight the need for future research to focus on the specific circumstances in which health events result in (social) learning effects and their heterogeneity. Health prevention efforts will remain a key priority in societies faced with an increasing and unequally distributed burden of disease from preventable causes. Understanding how individuals learn about their own health and how perceptions influence behaviours could help in addressing these future challenges efficiently through tailored information campaigns or interventions.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1-19.

Au, N., & Johnston, D. W. (2014). Self-assessed health: what does it mean and what does it hide?. *Social Science & Medicine*, 121, 21-28.

Arni, P., Dragone, D., Goette, L., & Ziebarth, N. R. (2021). Biased health perceptions and risky health behaviours—Theory and evidence. *Journal of Health Economics*, 76, 102425.

Bago d'Uva, T., O'Donnell, O., & van Doorslaer, E. (2020). Who can predict their own demise? Heterogeneity in the accuracy and value of longevity expectations. *The Journal of the Economics of Ageing*, 17, 100135.

Baji, P., & Bíró, A. (2018). Adaptation or recovery after health shocks? Evidence using subjective and objective health measures. *Health Economics*, 27(5), 850-864.

Banks, J., Kesternich, I., & Smith, J. P. (2021). International differences in interspousal health correlations. *Health Economics*, 30(5), 1152-1177.

Bernheim, B. D., & Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 94(5), 1558-1590.

Binno, S. (2016). 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *European Heart Journal*, 37, 2315-2381.

Biró, A. (2016). Differences between subjective and predicted survival probabilities and their relation to preventive care use. *The BE Journal of Economic Analysis & Policy*, 16(2), 807-835.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5), 803-843.

Bouckaert, N., Gielen, A. C., & Van Ourti, T. (2020). It runs in the family–Influenza vaccination and spillover effects. *Journal of Health Economics*, 74, 102386.

Buxbaum, J. D., Chernew, M. E., Fendrick, A. M., & Cutler, D. M. (2020). Contributions Of Public Health, Pharmaceuticals, And Other Medical Care To US Life Expectancy Changes, 1990-2015: Study examines the conditions most responsible for changing US life expectancy and how public health, pharmaceuticals, other medical care, and other factors may have contributed to the changes. *Health Affairs*, 39(9), 1546-1556.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200-230.

CBS (2015) Gezondheidsmonitor 2012. Opbouw en instructie totaalbestad. https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/gezondheidsmonitor-2012-opbouw-en-instructie-totaalbestand

CBS (2017) Documentatie Gezondheidsmonitor Volwassen en Ouderen 2016. https://www.volksgezondheidenzorg.info/sites/default/files/gemon.pdf

Chandra, A., & Staiger, D. O. (2007). Productivity spillovers in health care: evidence from the treatment of heart attacks. *Journal of political Economy*, 115(1), 103-140.

Choudhry, N. K., Avorn, J., Glynn, R. J., Antman, E. M., Schneeweiss, S., Toscano, M., ... & Shrank, W. H. (2011). Full coverage for preventive medications after myocardial infarction. *New England Journal of Medicine*, 365(22), 2088-2097.

Choung, Y., Chatterjee, S., & Pak, T. Y. (2022). Depression and financial planning horizon. *Journal of Behavioural and Experimental Economics*, 98, 101877.

Cropper, M. L. (1977). Health, investment in health, and occupational choice. *Journal of Political Economy*, 85(6), 1273-1294.

Cubí-Mollá, P., Jofre-Bonet, M., & Serra-Sastre, V. (2017). Adaptation to health states: Sick yet better off?. *Health Economics*, 26(12), 1826-1843.

Cawley, J., & Ruhm, C. J. (2011). The economics of risky health behaviours. In *Handbook of Health Economics* (Vol. 2, pp. 95-199). Elsevier.

Darden, M., & Gilleskie, D. (2016). The effects of parental health shocks on adult offspring smoking behaviour and self-assessed health. *Health Economics*, 25(8), 939-954.

Darden, M. (2017). Smoking, expectations, and health: a dynamic stochastic model of lifetime smoking behaviour. *Journal of Political Economy*, 125(5), 1465-1522.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2), 315-72.

Dwyer, D. S., & Mitchell, O. S. (1999). Health problems as determinants of retirement: Are self-rated measures endogenous?. *Journal of health Economics*, 18(2), 173-193.

Doyle, J. J. (2011). Returns to local-area health care spending: evidence from health shocks to patients far from home. *American Economic Journal: Applied Economics*, 3(3), 221-43.

Fadlon, I., & Nielsen, T. H. (2019). Family health behaviours. *American Economic Review*, 109(9), 3162-91.

Feigin, V. L., Norrving, B., & Mensah, G. A. (2017). Global burden of stroke. *Circulation Research*, 120(3), 439-448.

Finkelstein, A., & McGarry, K. (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review*, 96(4), 938-958.

Freyaldenhoven, S., Hansen, C., Pérez, J. P., & Shapiro, J. M. (2021). Visualization, identification, and estimation in the linear panel event-study design (No. w29170). National Bureau of Economic Research.

Gan, L., Gong, G., Hurd, M., & McFadden, D. (2015). Subjective mortality risk and bequests. *Journal of Econometrics*, 188(2), 514-525.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254-277.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, 19(3), 403-420.

Grossman, M. (1972). The demand for health: a theoretical and empirical investigation. NBER Occasional Papers, 119.

Handel, B. R., Kolstad, J. T., Minten, T., & Spinnewijn, J. (2020). The social determinants of choice quality: evidence from health insurance in the Netherlands (No. w27785). National Bureau of Economic Research.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4), 605-654.

Hoagland, A. (2022). An ounce of prevention or a pound of cure. The value of health risk information. Working Paper.

Hodor, M. (2021). Family health spillovers: evidence from the RAND health insurance experiment. *Journal of Health Economics*, 79, 102505.

de Hond, A., Bakx, P., & Versteegh, M. (2019). Can time heal all wounds? An empirical assessment of adaptation to functional limitations in an older population. *Social Science & Medicine*, 222, 180-187.

Jürges, H. (2007). True health vs response styles: exploring cross-country differences in self-reported health. *Health Economics*, 16(2), 163-178.

Kaestner, R., Darden, M., & Lakdawalla, D. (2014). Are investments in disease prevention complements? The case of statins and health behaviours. *Journal of health Economics*, 36, 151-163.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959-976.

Lindeboom, M., & Van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 23(6), 1083-1099.

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social Science & Medicine*, 55(12), 2149--2158.

Meier, A. N. (2022). Emotions and risk attitudes. *American Economic Journal: Applied Economics*, 14(3), 527-558.

Nie, P., Wang, L., & Sousa-Poza, A. (2021). Health perception biases and risky health behaviours in China. *Applied Economics Letters*, 1-8.

Nielsen, T. H. (2016). The relationship between self-rated health and hospital records. *Health Economics*, 25(4), 497-512.

Pandya, A., Gaziano, T. A., Weinstein, M. C., & Cutler, D. (2013). More Americans living longer with cardiovascular disease will increase costs while lowering quality of life. *Health Affairs*, 32(10), 1706-1714.

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., ... & GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *Journal of the American College of Cardiology*, 76(25), 2982-3021.

Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1), 101-122.

Spitzer, S., & Shaikh, M. (2022). Health misperception and healthcare utilisation among older Europeans. *The Journal of the Economics of Ageing*, 22, 100383.

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of clinical epidemiology*, 66(8), S84-S90.

Thomas, L. R., and Mentzakis E. (2022). The direct spillover effects of diabetes diagnosis on lifestyle behaviours. Working Paper.

5

CHAPTER 5

Is the grass greener on the other side? – The Health Impact of Providing Informal Care in the UK and the Netherlands

Based on: Bom, J., & Stöckel, J. (2021). Is the grass greener on the other side? The health impact of providing informal care in the UK and the Netherlands. *Social Science & Medicine*, 269, 113562.

Abstract:

Facing rapidly ageing populations, many Western countries aim to stimulate informal care provision as a way to meet the growing long-term care (LTC) demand. While various studies report the impact of providing informal care on the health of caregivers, it is less clear whether and to what extent this impact differs across countries. Using propensity score matching we match caregivers to similar non-caregiving individuals using four waves of the Dutch Study on Transitions in Employment, Ability and Motivation and the UK Household Longitudinal Study. The samples consist of 8,129 Dutch and 7186 UK respondents, among which respectively 1,711 and 1,713 individuals are identified as caregivers. We explore whether the health impact of providing informal care differs by country once similar caregivers, in terms of the intensity of provided care, are compared. In both countries we find negative mental health effects of providing informal care. While these effects slightly differ by country, the main differences arise between subgroups of caregivers. Individuals that provide more than 20 hours of informal care per week, and those who face a double burden of care and full-time employment experience the most severe negative mental health effects. These results indicate that health effects of providing informal care are mediated by the specific caregiving context, allowing policymakers to use information on this context to provide targeted aid. In addition, it suggests that previously reported differences of caregiving effects across countries could be driven by differences in the population of informal caregivers which are shaped by countries' LTC policies.

1 Introduction

Facing rapidly ageing populations, many Western countries search for ways to meet the growing long-term care (LTC) demand. Informal care, care provided by friends and family members, is one of the ways in which this demand can be (partially) met while limiting direct monetary costs. Reliance on informal care, however, is not without disadvantages. Next to its potential impact on caregivers' labor market participation, various studies indicate that the provision of informal care negatively affects informal caregivers' health (see Bom et al., 2019a for a review). These health effects are not the same for all caregivers. Individual and contextual elements like age, the intensity of care provided and other responsibilities like formal employment may affect the care burden (e.g., Pearlin et al. 1990, Pinquart & Sorensen, 2011). At the same time, country-level factors like welfare state generosity and cultural norms might also influence the impact of caregiving as they shape the societal environment in which informal care is provided (Brandt, 2013).

The country specific context can affect the health impact of providing informal care in various ways. First, country specific elements could influence the type and intensity of provided care (Brandt, 2013). The *specialization theory* hypothesizes that in countries with a generous welfare state, division of labor between formal and informal caregivers is higher (Motel-Klingebiel, Römer & Von Kondratowitz, 2005; Igel et al. 2009). While formal LTC professionals provide intensive, highly skilled care services, informal caregivers can dedicate themselves to lower intensity care activities. In less generous welfare states, on the contrary, family members are required to provide highly intensive care themselves. Accordingly, this implies that the generosity of LTC schemes directly shapes the population of informal caregivers, both regarding the care intensity as well as who provides care in the light of other obligations, such as childcare or paid work. There is evidence for the LTC system generosity directly influencing the population characteristics of informal caregivers (Bakx et al., 2015). Differences in the composition of the caregiver population could lead to differences in the average and aggregate health impact of relying on informal caregivers as highly intensive and specialized care tasks are often more stressful for caregivers (Pearlin et al., 1990).

A second way in which the country context might affect the relationship between informal care and caregivers' health points towards social norms and expectations about the family's role in meeting care demand. In countries where caregiving is considered a duty of family-members, informal caregivers might feel more pressured to provide care (Verbakel, 2014). This role-captivity, the feeling of being obliged to provide care, might influence the impact of care tasks on well-being (Pearlin et al., 1990). Additionally, the availability of formal care could influence one's experience of informal care. Knowing that formal care would be available if needed might affect the perceived control and hence the ability to deal with the situation (Wagner & Brandt, 2018). Furthermore, countries could differ in the depth and efforts to identify and help caregivers in need of support, which might affect the experience of care provision as well (Kaschowitz & Brandt, 2017).

Some studies investigated the relation between informal caregiving and health from a cross-country perspective using cross-sectional data. Dujardin et al. (2011) for example compared the health differences between informal caregivers and non-caregivers in the UK and Belgium using census data and found that high intensity British caregivers have better health than their Belgian counterparts.

More recently, various studies used the European SHARE panel dataset to estimate the cross-country health effect of informal caregiving with differing conclusions. Brenna & Di Novi (2016) estimated the effect of maternal informal care on caregiving daughters' health using propensity score matching methods to address endogeneity concerns. Their results indicate a North-South-gradient in the mental health effects of caregiving with negative effects only occurring within the context of Southern-European countries where LTC schemes provide little public support. Uccheddu et al. (2019) studied transitions into and out of spousal caregiving in Europe using fixed-effects models and again found that health effects are strongest in Southern and Eastern European countries. Kaschowitz & Brandt (2017) use the same dataset in combination with comparable panel data from the UK to estimate a set of fixed-effects models. Contrary to the results mentioned above, they find that caregiving negatively affects mental health across most European countries, irrespective of the specific policy context. Instead, the caregiving context, whether care is provided inside or outside the household, and the likely associated differences in care intensity seem to be the main determinant of the size of the observed differences. However, as informal care intensity is not captured in the SHARE data Kaschowitz & Brandt (2017) cannot test this hypothesis. Lastly, Van den Broek & Grundy (2018) studied the difference between caregiving effects in Sweden and Denmark by using a difference-in-differences approach to explore the impact reduced formal LTC availability in Denmark. Using the respective SHARE country samples, their results indicate that the reduced LTC availability led to lower quality of life among Danish caregivers. Hence, they conclude that LTC coverage directly shapes the impact of caregiving on caregivers' mental health, however again unavailable information on care intensity obstructs a more in-depth analysis that would allow policymakers to identify those groups of caregivers most vulnerable to insufficient LTC coverage.

A limitation of the current studies is hence that they incorporate little information on the caregiving intensity. As a result, they cannot determine whether the observed average differences in health effects are driven by differences in the caregiving population (e.g., a higher share of high-intensity caregivers) or whether other country differences (like support options and social norms) play a role as well. The current study combines two independent panel datasets from the Netherlands and the UK which contain detailed information on the caregiving context. This allows us to explore whether caregiving effects differ by country once compared at similar intensity levels. Additionally, similarity and size of both datasets facilitates a two-country comparison, instead of grouping several countries with different long-term care systems together. Lastly, we focus on starting informal caregivers. This allows us to measure the causal impact of becoming an informal caregiver on individuals' health without the potential bias that results from jointly analyzing longer-term and starting caregivers. We therefore contribute to the ongoing debate on the cross-country differences in caregiving effects by disentangling these effects at the intensive and extensive margin, an important distinction for LTC policymakers.

2 Background

To study the difference in caregiving effects between the Netherlands and the UK it is important to understand the differences in their LTC systems. Table 1 provides an overview of their LTC systems. Both countries are relatively similar in terms of the share of (dependent) elderly within the population, with slightly more elderly in the Netherlands. However, they differ strongly in terms of LTC expenditures.

Generosity of LTC systems

The difference in public LTC-spending reflects the generosity of their LTC systems. The Netherlands has a universal and comprehensive LTC system, irrespective of age or income, everyone requiring care is entitled to the benefits of this scheme (Mot, 2010). The system is largely publicly funded, copayments contribute only a small fraction (Maarse & Jeurissen, 2016). These copayments depend on type and duration of care, age, household composition, income, and (as of 2013) wealth. The payments are capped and cannot exceed the household income (Bakx et al., 2020a).

In the UK LTC is organized in a mixed-system combining universal and means-tested benefits. Health services and health related LTC components, such as nursing care, are provided for free by the National Health Service (NHS) (Colombo et al., 2011). Home care, day care and nursing home care are the responsibility of local authorities (Glendinning, 2013). This care is offered via a safety-net structure requiring users to deplete their wealth before obtaining publicly funded care (Colombo et al., 2011). Individual income and assets determine whether a service is (partly) covered (NHS, 2018a). Currently only individuals with assets below GBP 14,250 (approximately €16,886) will receive full-coverage (NHS, 2018b).

Role of informal carers in the system and available support

In both countries informal care is common: about 17-18% of the 50+ population identified him/herself as an informal caregiver (OECD, 2019). However, the average time spent caring strongly differs. According to the European Social Survey of 2014, 17% of the UK caregivers provides more than 20 hours of care per week compared to 8% in the Netherlands. This higher number of intensive informal caregivers in the UK seems to reflect the country's strong reliance on informal caregivers, which can be a result of only publicly funding non-health related formal LTC in case of low income/wealth and primarily directing formal care at people who do not receive informal care (Comas-Herrera et al., 2010). Both countries offer a wide range of support to informal caregivers and are among the few countries that have a national policy that targets this group. Furthermore, both countries offer (under different regulations) financial support, respite care, training, and counseling for caregivers (Courtin et al., 2014).

Hypotheses concerning the impact of informal care

Facing a different context, we formulate the following hypotheses regarding the differences in caregiving effects between the Netherlands and the UK: (i) As it offers more generous formal

LTC compared to the UK we expect caregivers in the Netherlands to 'specialize' and provide more low intensity care which might have a lower impact on their health. (ii) Additionally, norms and caregiving support might affect the relation between informal care provision and health, irrespective of the type of care that is provided. Whereas both countries offer a relatively comparable level of support for caregivers, norms potentially differ. Responses to the Eurobarometer (2007) for example indicate that UK citizens more often consider elderly care the responsibility of the family. More recent studies, although not using identical questions, also indicate different norms between the UK and the Netherlands. In 2016, 9% of the UK population stated that care to older individuals should be primarily provided by family and friends (British Social Attitudes Survey, 2016), while only 4% of Dutch respondents indicated that care for a dependent parent was predominantly a task for the family (SCP, 2016). These different attitudes might make caregivers in the UK feel more pressured towards providing care. Therefore, we expect a larger caregiving burden among UK caregivers compared to similar Dutch caregivers.

Table 1: Key figures regarding the LTC systems in the Netherlands and the UK

	Netherlands	**United Kingdom**
Share of population aged 65+[1]	18.7%	18.1%
Share of 65+ reporting some/severe limitations in daily activities[2]	47.9%	44.9%
LTC expenditures (health and social components) governmental and compulsory schemes (% GDP)[3]	3.7%	1.4%
Beds in LTC facilities per 1000 65+ inhabitants[4]	74.8	45.6
% informal caregivers among 50+ population[5]	16.8%	18.2%
% of caregivers providing at least 20 hours of care per week[6]	8%	17%
Care services available to informal caregivers[7]	Carers and care receivers' allowance, additional benefits, paid leave, unpaid leave, flexible work arrangements, training/education, respite care, counseling	Carers and care receivers' allowance, additional benefits, unpaid leave (for couple days under emergency situations), flexible work arrangements, training/education, respite care, counseling

Data concerning 2017 from 1. OECD (2020b); 2. OECD (2019); 3. OECD (2020a) and 4. OECD (2020c). 5. Data for 2017 from SHARE/ELSA from OECD (2019), UK in this case refers to England. 6. European Social Survey (2014). 7.Data from Colombo et al. (2011) for 2010.

3 Methods

It is not possible to study the impact of informal care provision on health by comparing the health of caregivers and non-caregivers as certain individuals, for example those with lower health, might be more likely to provide informal care. To account for these selection effects, we use propensity score matching. Following Schmitz & Westphal (2015) we construct a score of someone's propensity of providing informal care. This propensity score of informal caregiving is based on various elements that might affect the caregiving decision. The variables included can be grouped into three categories. The first, care obligations, covers information on parents and spouses to capture the presence of individuals in potential need of care. We further include whether both parents are alive and whether siblings are present to capture alternative informal care sources. The second category contains information on respondents themselves such as personal characteristics (age, sex), socio-economic status (marital and employment status, household income) and household structure. The third category contains information on individuals' health status using the MCS and PCS values and the self-reported presence of long-standing illnesses/disability. The complete list of variables used can be found in Table 2 as well as Tables O5.A1 to O5.A3 in the Online Supplement.

By matching caregivers and non-caregivers based on their propensity of providing informal care we assume that the remaining difference in health is due to caregiving. Or phrased differently, in absence of informal care provision the health of caregivers and matched non-caregivers would be similar and differences are causally attributable to informal caregiving. This assumption of conditional independence is the main assumption underlying our estimation strategy. To make it more credible, we follow Lechner (2009) and match upon control variables reported in the year before caregiving starts as the previous caregiving status captures most unobserved heterogeneity and to ensure that informal caregiving cannot affect the covariates.

The propensity scores are calculated using probit models that estimate the propensity of starting informal care provision conditional upon all variables potentially affecting the care decision in the preceding wave. We separately estimate these propensity scores for the Dutch and UK sample using the same approach and covariates.

After estimating the propensity scores, we match starting caregivers to non-caregivers using a kernel matching approach. We make use of the Stata command psmatch2 (Leuven & Sianesi, 2003) using an Epanechnikov kernel with a bandwidth of 0.03. Using alternative specifications with higher and lower bandwidth values (0.01 and 0.06) led to highly similar results (results available upon request). We regress informal care provision on health while adding all covariates from the pre-treatment wave. Adding the covariates to the regression next to matching based on the same covariates is referred to as double-robust. This corrects for remaining differences in covariates distributions between the two groups (Lechner, 2009; Rubin, 1979). With this analysis we estimate the average treatment effect on the treated (ATT). The ATT represents the mean difference in health between the group of informal caregivers (the treated) and the matched non-caregiving individuals.

We assess whether our matching strategy achieved its goal of balancing covariates using the standardized bias (Rosenbaum & Rubin, 1985). The standardized bias can be calculated for each covariate in the model by taking the difference in means between the treatment and control group and dividing it by the standard deviation of the control group.

4 Data

We use two similar datasets providing representative samples of the Dutch and UK population. The Study on Transitions in Employment, Ability and Motivation (STREAM) panel survey was carried out in the Netherlands. We include the first four waves of data annually collected via self-completion online surveys from 2010 to 2013 among the Dutch population aged 45-64 years. This sample is drawn from an existing internet panel (Ybema et al., 2014). For the UK we use the first four waves of the United Kingdom Household Longitudinal Survey, commonly known as Understanding Society (USoc; University of Essex, 2019). Data from the USoc is collected online or via face-to-face interviews among the 16+ population, data collection is annually and started 2009. Ethics approval has been obtained by the USoc and STREAM researchers and therefore no further ethical approval was required.

Informal care definition

We construct a binary variable indicating whether an individual provides informal care. In the Dutch survey informal caregivers are identified in case they positively answered to the following question and answer option: *'Did you in the past 12 months spend part of your time on any of the following activities?'* answer option: *'Giving Informal Care'*. In the UK sample individuals are identified as caregivers in case they affirmatively answered to at least one of the following two questions: '*Is there anyone living with you who is sick, disabled or elderly whom you look after or give special help to (for example a sick, disabled or elderly relative/husband/wife/friend etc.)?'* or *'Do you provide regular service or help for any sick, disabled or elderly person not living with you? [Exclude help provided in course of employment]'*. Both studies furthermore ask for care intensity, the average number of hours someone provides informal care per week. For both we construct dummy variables indicating low intensity (less than 10 hours of care per week), medium intensity (between 10 – 20 hours of care per week) and high intensity caregivers (more than 20 hours of care per week).

Health outcomes

To capture the health effect of informal caregiving we use the 12-item Short Form Health Surveys (SF-12). This health survey consists of 12 self-reported questions related to health in the past four weeks. Based on these questions the Physical Component Summary (PCS) and the Mental Component Summary (MCS) can be derived, relating to physical and mental health. Both scales are validated and range from 0 (lowest health) to 100 (optimal health) and transformed to have a mean of 50 and a standard deviation of 10 (Ware, Kosinski, & Keller, 1995).

Other covariates

We estimate the individual's propensity of providing informal care based on a broad set of variables that might affect someone's caregiving decision and health status. Variables related to the health and demographics of the respondent are present in both datasets and in most cases easily comparable as (a) they use the same instruments (e.g. the MCS and PCS) or (b) because the questions are straightforward and highly similar in both countries (e.g. age or employment status of respondent). We however want to match on a broad set of variables that also contains information about the income and family structure of the respondent. This information is available in the USoc but not in STREAM. We therefore enrich the Dutch survey dataset with information from administrative sources covering information on: personal and household income from the tax authority and information about the family structure from the municipal register.

For the variables related to family structure we argue that the differences between the self-reported versus administrative data are minimal. Comparing self-reported and tax-registered income we however must be careful as self-reported income might suffer from reporting bias. However, we use the variables to predict informal care provision separately for both countries, hence no direct comparison between both values is needed. For our analysis we assume that any reporting bias in the income variable is stable throughout the income distribution of the respective country sample. An overview of the definition and source of all used variables is available in Table O5.A1 in the Online Supplement.

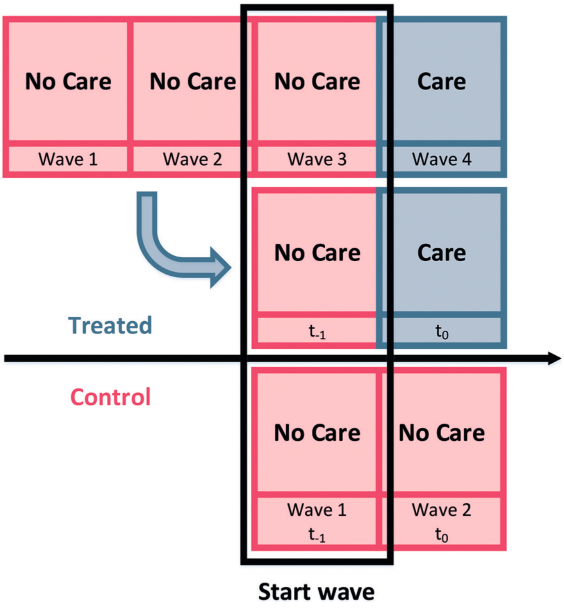
Sample selection

In order to make both datasets comparable we restrict the samples as follows: (i) We include respondents aged 45-65 in the first wave; (ii) we use information from the first four waves of the surveys ranging from 2009/2010 – 2013/2014. These selection criteria are motivated by the fact that the Dutch dataset only covers individuals aged 45-65 and we have access to its first four waves spanning 2010 to 2013, hence we use a similar subset of the USoc data. Further, (iii) we condition the datasets on availability of all control variables in the first wave and all needed outcome variables in the first and second wave; (iv) we exclude all individuals that already provided informal care in the first wave as we only look at starting caregivers. Eventually the samples consist of 8,141 Dutch and 7,187 UK respondents.

Time structure

For both datasets we define a relative time variable (t) whose value depends on an individual's first reported care-episode (see Figure 1 for a graphical representation). Within the control group t_{-1} is normalized to the individuals first appearance in the survey as these respondents do not report any care episode during their participation. Among caregivers t_{-1} is defined as the period before the first reported caregiving episode. For example, an individual entering the panel in 2010 and responding to the survey for four consecutive waves but only starting to provide informal care in wave 4 is included for two periods, t_{-1} (wave 3) to t_0 (wave 4). This time structure is chosen to maximize the number of informal caregivers that we can observe.

Figure 1: Visualization of the constructed time structure of the sample



Source: Own illustration.

5 Results

Descriptives & matching results

Table 2 provides an overview of the composition of caregivers in both datasets. These samples are constructed to maximize the number of starting caregivers and hence contain all individuals who started care provision in 2011-2013. There are slightly more starting caregivers in the UK where about 24% of the sample starts care provision compared to 21% of the Dutch sample. The share of female and parental caregivers is higher in the UK than in the Netherlands and a larger share of the caregivers in the UK provides medium or high intensity care compared to the Dutch sample. About half of the caregivers in both countries have a full-time job next to their caregiving duties.

Table 2: Descriptive statistics of both samples

	Dutch Sample		UK Sample	
Starting caregiver (% of sample)	1,711	(21.1%)	1,713	(23.8%)
Female caregivers (% of caregivers)	865	(50.6%)	1,019	(58.5%)
Low Intensity caregivers	1,378	(80.5%)	1,302	(76.0%)
Medium Intensity caregivers	181	(10.6%)	205	(12.0%)
High Intensity caregivers	135	(7.9%)	177	(10.3%)
Spousal caregivers[1]	203	(12.0%)	214	(12.0%)
Parental caregivers[1]	688	(40.0%)	954	(56.0%)
Full-time employed (% of sample)	4,634	(57.0%)	3,863	(53.8%)
Full-time employed & caregiver (% of caregivers)	856	(50.0%)	844	(49.3%)
Individuals	**8,129**		**7,186**	

1. Dutch values are imputed as information on care recipient for the Dutch data is only available in 2012-2013.
Sources: STREAM Wave 1-4 & USoc Wave 1-4.

To match caregivers and non-caregivers we estimate propensity scores of providing informal care. Table 3 provides an overview of the propensity score estimations. In both samples especially the variables related to care obligations are strongly correlated to someone's propensity of providing informal care. As parents tend to provide care to each other, the presence of both parents is negatively associated with informal care. The age of these parents, which acts as a crude proxy for the rate of dependency of the individual, increases the informal care propensity. Furthermore, females and more highly educated individuals are more likely to provide care whereas the presence of young children is negatively related to informal care provision.

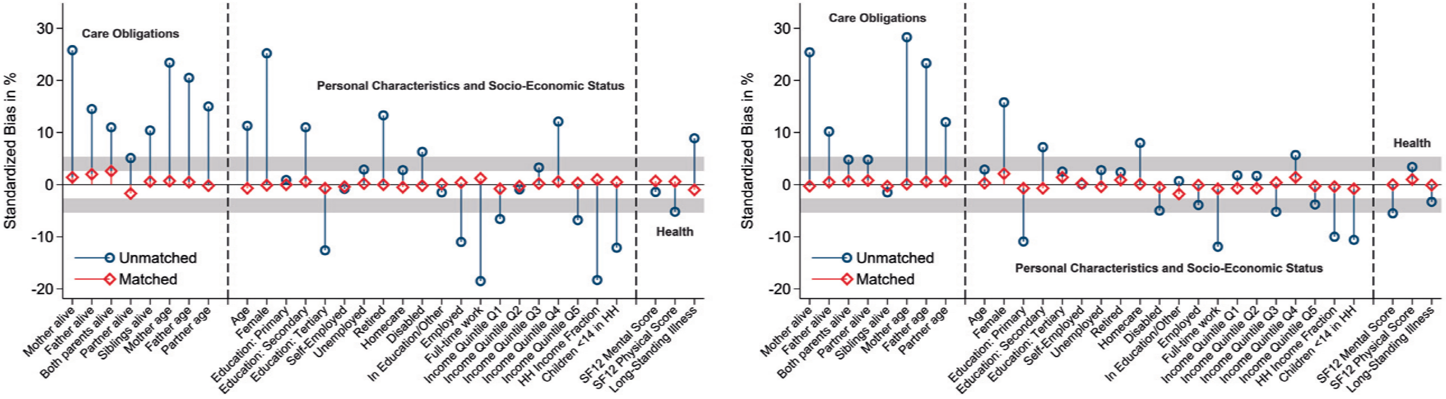
In both countries informal caregivers differ from non-caregivers. As can be seen in Online Supplement Tables O5.A2 and O5.A3 there is a strong imbalance between the individuals that started to provide care and those who did not do so. This imbalance is depicted in Figure 2 by plotting the pre-matching (blue) and post-matching (red) standardized bias values for each control included. Before the matching there is considerable imbalance between the non-caregiver and caregiver samples with many variables exceeding the 3-5% standardized bias threshold (grey bar). The matching succeeds in correcting this imbalance with the standardized bias between the matched control group and the treatment group falling below the thresholds for all considered variables. For all analyses we exclude respondents that were identified as off support, this equals to 1 in the UK and 12 in the Dutch sample.

Table 3: Propensity score estimates

	Dutch sample		**UK sample**	
	Coefficient	Standard Error	Coefficient	Standard Error
Care Obligations				
Mother alive	0.425***	(0.044)	0.451***	(0.047)
Father alive	0.429***	(0.069)	0.428***	(0.071)
Both parents alive	-0.259**	(0.084)	-0.459***	(0.085)
Living partner	0.103*	(0.042)	0.063	(0.043)
Living siblings	-0.003	(0.051)	-0.061	(0.056)
Age of mother	0.023***	(0.005)	0.026***	(0.005)
Age of father	0.022***	(0.006)	0.023***	(0.006)
Age of partner	-0.000	(0.004)	0.008*	(0.003)
Personal Characteristics and Socio-Economic Status				
Age	0.014**	(0.005)	-0.003	(0.005)
Female	0.309***	(0.043)	0.113**	(0.039)
Secondary Education (*Ref. Primary education*)	0.154***	(0.041)	0.156***	(0.044)
Tertiary Education	0.232***	(0.045)	0.098*	(0.048)
Self-employed *(Ref. Employed)*	-0.138	(0.064)	0.036	(0.057)
Unemployed	0.118	(0.104)	0.064	(0.087)
Retired	0.396***	(0.094)	-0.016	(0.065)
Homecarer	0.109	(0.153)	0.124	(0.093)
Disabled	0.121	(0.084)	-0.167	(0.093)
Studying or other activities	-0.273	(0.437)	0.003	(0.260)
Working Full-Time	-0.071	(0.047)	-0.167**	(0.049)
Income quintile 2 (Ref. Income quintile 1)	-0.007	(0.050)	0.066	(0.053)
Income quintile 3	-0.038	(0.052)	0.066	(0.055)
Income quintile 4	0.054	(0.054)	-0.051	(0.059)
Income quintile 5	0.158**	(0.057)	0.072	(0.060)
HH Income Fraction	-0.066	(0.084)	-0.034	(0.070)
Children<14 in household	-0.097	(0.057)	-0.146**	(0.052)
Health				
SF12 - Mental Score	-0.001	(0.002)	-0.006**	(0.002)
SF12 - Physical Score	0.001	0.002	0.000	(0.002)
Longstanding illness	0.107**	(0.038)	-0.033	(0.040)
Observations	**8,141**			**7,187**
Pseudo R^2		0.06		0.05

*p<0.05, **p<0.01, *** p<0.001, standard errors in parentheses. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

Figure 2: Standardized bias before and after matching.



(a) Dutch sample

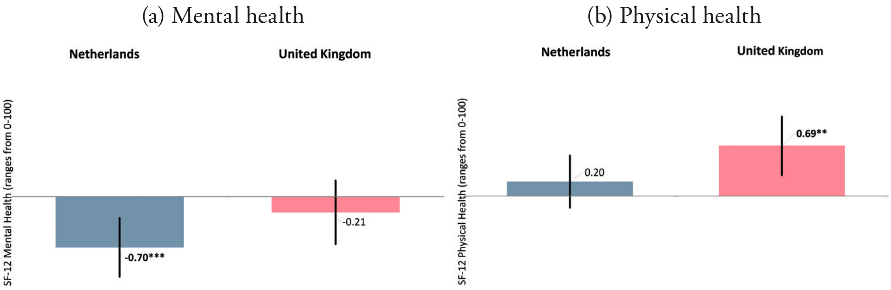
(b) UK sample

Note: The grey lines mark the standardized bias range of 3-5%, following the rule of thumb suggested by Caliendo & Kopeinig (2008) we consider balance to be sufficient when the bias is below 3-5%. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

Main results

Our baseline analysis estimates the impact of any informal care provision on health. Throughout the main text we present our results graphically, results can be found in Tables O5.A4 to O5.A6 in the Online Supplement. Figure 3 presents the impact of any care provision on (a) mental and (b) physical health. The bar presents the ATT, the confidence intervals are depicted at 95%. Dutch caregivers experience a direct negative mental health effect of -0.70 (p<0.001), whereas the negative impact of care provision in the UK is considerably smaller and insignificant. For physical health, on the contrary, informal care provision has a positive effect of 0.69 (p<0.01) in the UK whereas no significant impact is present among the Dutch caregivers.

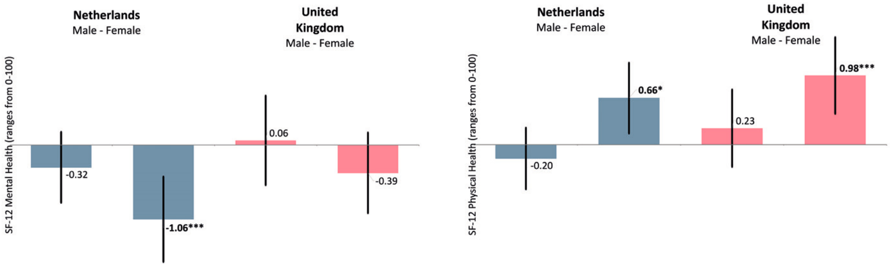
Figure 3: Impact of any informal caregiving on mental/physical health


(a) Mental health
(b) Physical health

Confidence intervals at 95%, *p<0.05, **p<0.01, *** p<0.001. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

As some studies indicate potential differences in the caregiving effect between males and females, we separately estimate the caregiving effect for both genders. When stratifying Dutch caregivers by gender we find a negative mental health effect of -1.06 (p<0.001) for females and no significant impact on males. In the UK, the mental health impact of care provision is larger for females than for males although both estimates are insignificant at a 95% level. Turning to physical health, we observe a different pattern. In both countries any informal care provision has a positive effect on the physical health of female caregivers. For male caregivers no effects are found in the physical health domain in either of the two countries.

Figure 4: Impact of any informal care on mental/physical health by gender



(a) Mental health

(b) Physical health

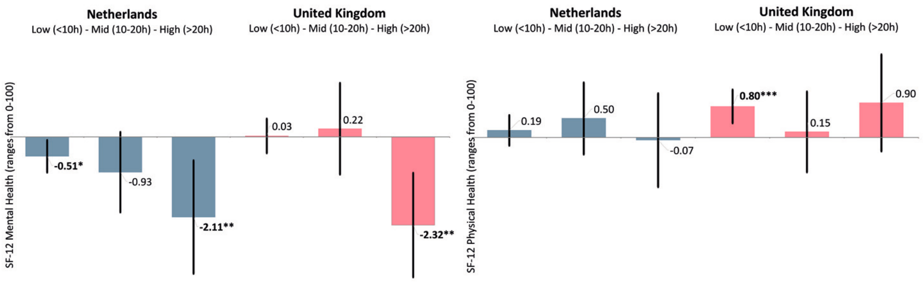
Confidence intervals at 95%, *p<0.05, **p<0.01, *** p<0.001. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

By comparing the health impact of any care provision between both countries we ignore underlying differences in the composition of the caregiver population. This composition might however differ per country, for example as a result of differences in the LTC system. As presented in Table 2, there are for example slightly more medium and high intensity caregivers in the UK. To compare similar caregivers, we construct three groups based on the hours of care provided.

Figure 5 presents the treatment effects when separately estimated for low, medium and high intensity caregivers. The figure clearly depicts that the impact of care provision strongly differs by the amount of care provided. In both countries, high intensity caregivers experience the largest mental health effects. This health impact of providing more than 20 hours of informal care per week is similar in both countries with an impact of -2.11 (p<0.01) on the MCS in the Netherlands compared to -2.32 (p<0.01) in the UK. The pattern of the impact of informal care provision by care intensity however slightly differs between the two countries. In the Netherlands, a clear dose-response relationship is visible; all caregivers experience negative mental health effects that grow in response to care intensity. In the UK, low and medium intensity care providers are not affected, only high intensity caregivers experience a strong decline in their mental health.

The intensity-patterns also differ when focusing on physical health. In the Netherlands, no health effects are present when separating the sample by care intensity. In the UK, an initial positive physical health effect is present for low intensity caregivers and absent for medium intensity caregivers. For individuals providing more than 20 hours of care per week the estimates seem to point again to a positive effect although the results are insignificant.

Figure 5: Impact of informal caregiving on mental/physical health by care intensity


(a) Mental health

(b) Physical health

Confidence intervals at 95%, *p<0.05, **p<0.01, *** p<0.001. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

Next to the intensity of care provided, other contextual elements could influence the care burden. Individuals might for example experience increased caregiving strain when providing informal care in combination to full-time employment. Facing multiple responsibilities might namely lead to cross-pressures, like fatigue or dissatisfaction about decreased productivity at work due to caregiving tasks (Pearlin et al., 1990). Using German data, Schmitz & Stroka (2013) found that individuals experiencing a double burden of care and work were more likely to use antidepressant drugs and tranquillizers. Again, country differences in terms of available alternatives and norms and support could make this situation more prevalent or straining.

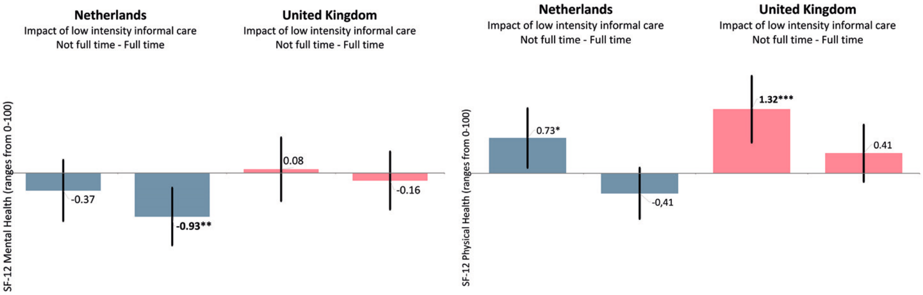
To estimate the impact of care provision for individuals experiencing a double burden we compare the health impact of care provision between individuals in full-time employment to those not working full-time. We solely focus on individuals with stable workforce participation to exclude individuals that overcome the double burden of care and work by cutting down on working hours. In our samples this relates to excluding 7% (Netherlands) to 10% (UK) of our sample as these individuals experience changes in their work participation (from full-time to no work/part-time and vice versa). Starting informal caregivers seem slightly more likely to adjust their work participation than the control group of non-caregivers. In the UK, 10.7% of the starting caregivers change work participation compared to 9.6% in the control group. In the Netherlands these numbers equal 7.6% and 6.8%. Additionally, due to sample size limitations we solely compare employment status by individuals providing either low or medium to high intensity care provision (>10h of care per week).

Figure 6 shows that the experienced mental health effect of providing low intensity informal care slightly differs by employment status. In the Netherlands, the estimates of the caregiving effect point in the negative direction for all low intensity caregivers. The impact is however larger and significant (-0.93, p<0.01) among full-time workers. In the UK no mental health effects are present for either of the two groups. With regards to physical health effects we observe larger differences between full-time working individuals and those who work less hours or not at all. In both countries, the latter group experiences a positive physical health effect of caregiving of respectively 0.73 (p<0.05) and 1.32 (p<0.001). This effect on physical health is absent or even negative among full-time working individuals.

A double burden might especially be present for individuals who next to a full-time job provide many hours of informal care. Figure 7 shows that in both countries indeed the mental health effect of providing more than 10 hours of informal care per week is larger for individuals working full-time. Interestingly the mental health effect of medium or high-intensity care even becomes insignificant among individuals who do not work full-time. For physical health, the initial positive effects disappear when focusing on intensive informal care.

Figure 6: Impact of low informal caregiving on mental/physical health by employment

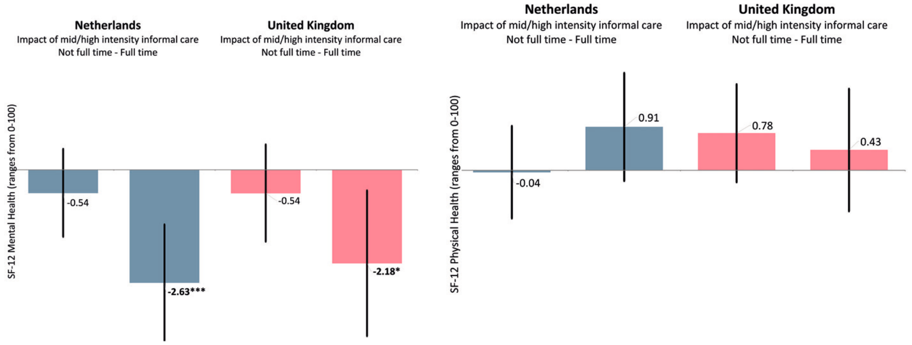
(a) Mental health (b) Physical health



Confidence intervals at 95%, *p<0.05, **p<0.01, *** p<0.001. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

Figure 7: Impact of mid/high intensity informal caregiving on mental/physical health by employment

(a) Mental health (b) Physical health



Confidence intervals at 95%, *p<0.05, **p<0.01, *** p<0.001. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

6 Robustness checks

To assess the robustness of our results we perform various robustness checks First, we assess whether our results are robust to our choices in the matching strategy. We check whether our results are driven by extreme propensity scores by excluding the highest/lowest 5%. Additionally, we check whether our results differ when re-estimating the propensity scores for the intensity groups separately. The results are highly comparable to those presented in the main specification (detailed results are available upon request).

Second, we test how sensitive our results are with regards to a violation of the main identifying assumption of conditional independence. We follow Ichino et al. (2008) who propose a simulation-based sensitivity analysis for propensity-score based treatment effects to unobserved variables that should have been included in the propensity score estimation. In the given context such an unobserved variable might be personality characteristics or norm perceptions that would influence an individual's likelihood to provide care (selection effect *s*) and their mental health in absence of providing care (outcome effect *d*). Another motivation for such a sensitivity analysis is the fact that we only observe caregivers before care provision (t_{-1}) and when they report to have started caregiving (t_0). We therefore cannot observe the actual moment of caregiving onset while additionally the actual cause for caregiving onset might have a direct effect (e.g. a health shock to a family member).

The method simulates a confounder with specific values for *s* and *d* in order to assess the sensitivity of estimated treatment effects to the inclusion of such unobserved confounders. Table 4 depicts our estimated treatment effects when simulating a confounder that combines the strongest observed selection and outcome effects. As the estimated values illustrate, our results are robust to such a simulated confounder. Table O5.A7 in the Online Supplement depicts the estimated selection and outcome effects for all our covariates.

Table 4: Mental Health Effects of High Intensity Caregiving - Sensitivity Analysis

	Dutch Sample		**UK Sample**	
	MCS	**PCS**	**MCS**	**PCS**
Confounder with properties: s=-0.2 d=0.15	-2.716***	-1.109	-2.478***	0.408
	(0.849)	(0.743)	(0.789)	(0.779)
Confounder with properties: s=0.2 d=0.15	-2.875***	-1.103	-2.910***	0.073
	(0.843)	(0.728)	(0.784)	(0.757)
Confounder with properties: s=-0.2 d=0.45	-2.548***	-1.143	-2.149***	0.646
	(0.863)	(0.766)	(0.824)	(0.814)
Confounder with properties: s=0.2 d=0.45	-3.131***	-1.124	-3.519***	-0.418
	(0.868)	(0.739)	(0.789)	(0.761)
Control	6,418			5,473
Treatment	135			177

*p<0.05, **p<0.01, *** p<0.001, standard errors in parentheses. Sources: STREAM Wave 1-4 & USoc Wave 1-4.

7 Discussion & Conclusion

While several studies have reported negative health effects of informal care provision on the caregivers' health, there remains uncertainty with regards to their causal nature and the differences of these effects across countries and caregiver subgroups. Using a propensity score matching approach and two comparable panel-data sets, we estimated the health impact of providing care in the UK and the Netherlands. Doing so, we investigated whether observed average health differences between informal caregivers and non-caregivers within and across countries are attributable to the composition of the caregiver populations in each country.

First, our results highlight the link between the generosity of LTC systems and the hours of informal care provision. We hypothesized that the share of high intensity caregivers would be higher in the UK than in the Netherlands as the LTC system is less generous. In our samples this is indeed the case with slightly more caregivers providing more than 10 hours of weekly care in the UK compared to the Netherlands. However, it is noteworthy that our samples seem to understate the true differences that become apparent when looking at population wide estimates (ONS, 2013; SCP, 2016) or results from the ESS (2014) which indicate that the share of intensive caregivers is much higher in the UK.

Second, we show the importance of considering care-intensity when comparing average caregiving effects across countries. In both countries especially individuals providing more than 20 hours of weekly care experience large negative mental health effects. Individuals providing less intensive care do not experience any mental health effect (United Kingdom) or a similar negative but considerably smaller one (Netherlands). These findings are consistent with the findings of Kaschowitz & Brandt (2017) who hypothesized that differences between groups of caregivers (like care intensity) determine the average impact of caregiving.

Third, our results uncover interesting patterns regarding the physical health effects of caregiving. We observe small positive physical health effects among caregivers who provide less than 10 hours of care per week and (for the Dutch sample) do not work full-time. While these small positive effects disappear with increasing care intensity, they indicate that low intensity caregiving might lead to small increases in physical health, possibly due to increased physical activity. However, these results should be taken with a pinch of salt. Di Novi et al. (2016) point out that self-reported health measures are prone to bias as individuals might change their judgement by taking the care-recipients health as a reference point. Although our use of a multi-item physical health measure that emphasizes the ability to perform certain tasks should mitigate this concern, it cannot be ruled out.

Lastly, our analyses indicate that especially caregivers who combine caregiving with full-time employment experience large negative mental health effects. These results are estimated only on those individuals who are in stable full-time employment and hence exclude individuals who, potentially as a result of their caregiving tasks, changed their employment status. This could for example be the case when someone reduced working hours due to caregiving-related health issues. While this might introduce a downward bias and raise concerns about the external validity of our results, recent evidence suggests that there is no effect of informal caregiving on labour force participation in both countries (Heitmüller, 2007; Rellstab et al., 2020). In the UK, intensive caregiving might however affect the number of hours worked (Heitmüller & Inglis, 2007). As across countries the relationship between informal care and labour market outcomes is diverse (for a review see Bauer & Sousa-Poza, 2015), there is nonetheless a need for further research into the complex relationship between health, labour market outcomes and the long-term care system.

The bulk of our results suggests that once the focus is narrowed to specific subgroups of caregivers the effects of informal caregiving are similar despite large differences across country's LTC systems. However, some differences between both countries still emerge. Dutch low-intensity caregivers experience small negative mental health effects whereas this is not the case in the UK. In turn, only UK caregivers experience small positive health effects among low-intensity caregivers while in the Netherlands this only occurs among unemployed or part-time working individuals. Lastly, there is some suggestive evidence that women in the Netherlands experience a larger mental health burden. These differences can be driven by variation in support options or attitudes towards care provision, but also by differences in the type of care provided (e.g., personal care, household help) or the associated social norms. The country-specific differences in effect size are however considerably smaller compared to differences between subgroups of caregivers.

While our study provides important insights, there are several limitations that need to be taken into account when interpreting our results. First, we rely on a matching based empirical strategy aimed at estimating the causal effect of informal care on health. Matching estimators require the conditional independence assumption to hold and while we attempt to explore the impact of a potential violation of this assumption, an ideal empirical strategy would rely on plausible exogenous variation in informal caregiving status. As pointed out by Schmitz & Westphal (2015) many of the commonly applied instrumental variables in this literature, such as the presence of siblings or health shocks, are not without their own drawbacks. A more credible source of variation could be obtained from policy variation as recently done by Bakx et al. (2020b). However, given our cross-country perspective such an empirical strategy is not feasible.

A second limitation is that we cannot disentangle the caregiving effect, the impact of caring for someone, from the family effect, the impact of caring about someone (Bobinac et al., 2010). Possibly the observed mental health effects are driven by severe illness of a family member instead of the act of caring itself. Previous studies investigating the caregiving and family effect in the Netherlands (Bom et al., 2019b) and the UK (Stöckel & Bom, 2020) however found that inclusion of an indicator of the family effect does not affect the estimates of the caregiving effect. In addition, the absence of strong negative health effects among high-intensity caregivers who are not in full-time work seems to contradict that the effects we find are driven by the family effect. Importantly though this does not mean that there is no direct mental health effect associated with concerns regarding family members' health. Rather it seems likely that such an effect is already captured when conditioning on the mental health outcomes in the year prior to providing informal care.

Lastly, while the used datasets allow for similar matching, we still lack some important information that ideally should be considered. We rely on self-reported hours as our measure of care intensity. While this measure seems sufficient to capture the general differences between intensity levels it would be ideal to not only observe hours but also the specific tasks that were conducted as they are highly disease-specific and differ with regards to their perceived burden (Pearlin, 1999). In addition, we cannot observe the consumption of formal care.

Concluding, our results provide insights for both researchers and policy makers. First, they indicate that especially differences in caregiver characteristics drive the differences in observed health effects of care provision across countries. These insights can be used to specifically target support to those caregivers who experience the largest burden: those who provide most hours or care and those experiencing a double burden of care and full-time employment. Second, while we do not find large differences in health effects between both countries when comparing similar caregivers, this does not mean that country characteristics do not play a role. In countries with more generous LTC systems, and hence more formal care alternatives to informal care, less individuals seem to provide highly intensive care. As a result, less individuals experience severe health effects of care provision. It is important for policymakers to be aware of this relation between the coverage of LTC systems and the composition (and hence experienced health effects) of caregivers in order to make deliberate trade-offs between the aggregate costs of formal care versus the implications of informal care.

References

Bakx, P., De Meijer, C., Schut, F., & Van Doorslaer, E. (2015). Going formal or informal, who cares? The influence of public long-term care insurance. *Health Economics*, 24(6), 631-643.

Bakx, P., Bom, J., Tenand, M., & Wouterse, B. (2020a). Eigen bijdrage aan ouderenzorg. *Netspar Design Paper 13.*

Bakx, P., Wouterse, B., van Doorslaer, E., & Wong, A. (2020b). Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival. *Journal of Health Economics*, 73, 102354.

Bauer, J. M., & Sousa-Poza, A. (2015). Impacts of informal caregiving on caregiver employment, health, and family. *Journal of Population Ageing*, 8(3), 113-145.

Bobinac, A., Van Exel, N. J. A., Rutten, F. F., & Brouwer, W. B. (2010). Caring for and caring about: disentangling the caregiver effect and the family effect. *Journal of Health Economics*, 29(4), 549-556.

Bom, J., Bakx, P., Schut, F., & van Doorslaer, E. (2019a). The Impact of Informal Caregiving for Older Adults on the Health of Various Types of Caregivers. *The Gerontologist*, 59(5), e629-e642.

Bom, J., Bakx, P., Schut, F., & van Doorslaer, E. (2019b). Health effects of caring for and about parents and spouses. *The Journal of the Economics of Ageing*, 100196.

Brandt, M. (2013). INTERGENERATIONAL HELP AND PUBLIC ASSISTANCE IN EUROPE: A case of specialization? *European Societies*, 15(1), 26-56.

Brenna, E., & Di Novi, C. (2016). Is caring for older parents detrimental to women's mental health? the role of the european North-South gradient. *Review of Economics of the Household*, 14(4), 745-778.

van den Broek, T., & Grundy, E. (2018). Does long-term care coverage shape the impact of informal care-giving on quality of life? A difference-in-difference approach. *Ageing and Society*, 40 (6). pp. 1291-1308.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72.

Colombo, F., Llena-Nozal, A., Mercier, J., & Tjadens, F. (2011). Help wanted? PROVIDING AND PAYING FOR LONG-TERM CARE. OECD Health Policy Studies, OECD Publishing.

Comas-Herrera, A., Pickard, L., Wittenberg, R., Malley, J., & King, D. (2010). *European network of economic policy research institutes assessing needs of care in european nations THE LONG-TERM CARE SYSTEM FOR THE ELDERLY IN ENGLAND*

Courtin, E. Jemiai, N. & Mossialos, E (2014). Mapping support policies for informal carers across the European Union, *Health Policy*, 118(1): 84-94.

Di Novi, C., Jacobs, R., & Migheli, M. (2015). The quality of life of female informal caregivers: From Scandinavia to the Mediterranean Sea. *European Journal of Population*, 31, 309–333.

Dujardin, C., Farfan-Portet, M. I., Mitchell, R., Popham, F., Thomas, I., & Lorant, V. (2011). Does country influence the health burden of informal care? an international comparison between Belgium and Great Britain. *Social Science and Medicine*, 73(8), 1123-1132.

ESS Round 7: European Social Survey Round 7 Data (2014). Data file edition 2.2. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

Eurobarometer (2007). Health and long-term care in the European Union

Glendinning, C. (2013). Long term care reform in England: A long and unfinished story. In *Reforms in long-term care policies in Europe* (pp. 179-200). Springer, New York, NY.

Heitmueller, A. (2007). The chicken or the egg?: Endogeneity in labour market participation of informal carers in England. *Journal of Health Economics*, 26(3), 536-559.

Heitmueller, A., & Inglis, K. (2007). The earnings of informal carers: Wage differentials and opportunity costs. *Journal of Health Economics*, 26(4), 821-841.

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity?. *Journal of Applied Econometrics*, 23(3), 305-327.

Igel, C., Brandt, M., Haberkern, K. and Szydlik, M. (2009). Specialization between family and state. Intergenerational time transfers in western Europe, *Journal of Comparative Family Studies,* 40(2): 203-226.

Kaschowitz, J., & Brandt, M. (2017). Health effects of informal caregiving across Europe: A longitudinal approach. *Social Science and Medicine*, 173, 72-80.

Lechner, M. (2009). Sequential causal models for the evaluation of labor market programs. *Journal of Business & Economic Statistics*, 27(1), 71-83.

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Boston College Department of Economics, Statistical Software Components.

Maarse, J. A. M. H., & Jeurissen, P. P. P. (2016). The policy and politics of the 2015 long-term care reform in the Netherlands. *Health Policy*, 120(3), 241-245.

Mot E. (2010). The Dutch System of Long-Term Care, CPB Document, No. 204. Netherlands Bureau for Economic Policy Analysis, The Hague.

Motel-Klingebiel, A., Tesch-Römer, C. and von Kondratowitz, H.-J. (2005). Welfare states do not crowd out the family : evidence for mixed responsibility from comparative analyses. *Ageing & Society*, 25, 6, 863–882.

NatCen Social Research. (2017). British Social Attitudes Survey, 2016. [data collection]. UK Data Service. SN: 8252, http://doi.org/10.5255/UKDA-SN-8252-1

NHS (2018a). Financial assessment, https://www.nhs.uk/conditions/social-care-and-support-guide/help-from-social-services-and-charities/financial-assessment-means-test/

NHS (2018b).When the council might pay for your care, https://www.nhs.uk/conditions/social-care-and-support-guide/money-work-and-benefits/when-the-council-might-pay-for-your-care/

OECD (2019). Health at a Glance. https://doi.org/10.1787/4dd50c09-en

OECD (2020a). Health expenditure and financing. https://stats.oecd.org/Index.aspx?DataSetCode=SHA

OECD (2020b). Long term care. https://www.oecd.org/els/health-systems/long-term-care.htm

OECD (2020c). Long-term care resources and utilization: long-term care (health) recipients. https://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT#

ONS (2013). 2011 Census Analysis: Unpaid care in England and Wales, 2011 and comparison with 2001. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/articles/2011censusanalysisunpaidcareinenglandandwales2011andcomparisonwith2001/2013-02-15

Pearlin, L. I., Mullan, J. T., Semple, S. J., & Skaff, M. M. (1990). Caregiving and the stress process: An overview of concepts and their measures. *Gerontologist*, 30(5), 583-594.

Pinquart, M., & Sörensen, S. (2011). Spouses, adult children, and children-in-law as caregivers of older adults: A meta-analytic comparison. *Phychology and Aging*, 26(1), 1-14.

Rellstab, S., Bakx, P., Garcia-Gomez, P., & Van Doorslaer, E. (2020). The kids are alright-labour market effects of unexpected parental hospitalisations in the Netherlands. *Journal of Health Economics*, 69, 102275.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38. https://doi.org/10.2307/2683903

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366a), 318-328.

Schmitz, H., & Stroka, M. A. (2013). Health and the double burden of full-time work and informal care provision—Evidence from administrative data. *Labour Economics*, 24, 305-322.

Schmitz, H., & Westphal, M. (2015). Short- and medium-term effects of informal care provision on female caregivers' health. *Journal of Health Economics*, 42, 174-185.

SCP (2016). Voor elkaar? https://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2017/Voor_elkaar?utm_source=e-mailnieuwsbrief&utm_medium=email&utm_campaign=Nieuwe+publicatie+SCP

Stöckel, J. & Bom, J. (2020) The Dynamic Effects of Informal Caregiving on the Caregivers' Health. Netspar Discussion Paper 02/2020-011.

Uccheddu, D., Gauthier, A. H., Steverink, N., & Emery, T. (2019). The pains and reliefs of the transitions into and out of spousal caregiving. A cross-national comparison of the health consequences of caregiving by gender. *Social Science & Medicine*, 240, 112517.

University of Essex. Institute for Social and Economic Research. (2019). *Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. *12th Edition*. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-13

Verbakel, E. (2014). Informal caregiving and well-being in Europe: What can ease the negative consequences for caregivers? *Journal of European Social Policy*, 24(5), 424-441.

Wagner, M., & Brandt, M. (2018). Long-term care provision and the well-being of spousal caregivers: An analysis of 138 european regions. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 73(4), e24-e34.

Ware, J. E., Kosinski, M., & Keller, S. D. (1995). *SF-12: How to score the SF-12 physical and mental health summary scales* (2nd. ed). Boston, Massachusetts, The Health Institute, New England Medical Center

Ybema, J. F., Geuskens, G. A., Van Den Heuvel S. G., De Wind A., Leijten, F. R. M., Joling, C., et al. (2014). Study on transitions in employment, ability and motivation (STREAM): The design of a four-year longitudinal cohort study among 15,118 persons aged 45 to 64 years. *British Journal of Medicine & Medical Research*, 4(46), 1383-1399.

6

CHAPTER 6

Revisiting Longer-term Health Effects of Informal Caregiving: Evidence from the UK

Based on: Stöckel, J., & Bom, J. (2022). Revisiting longer-term health effects of informal caregiving: Evidence from the UK.
The Journal of the Economics of Ageing, *21*, 100343.

Abstract:

We estimate the longer-term and dynamic effects of providing informal care on caregivers' health in the United Kingdom. Using propensity score matching to address the endogeneity of informal care provision, we estimate static and sequential matching models exploring health effects at the extensive and intensive margin of informal caregiving and their persistence for up to five years. Our results suggest substantial negative health effects confined to the mental domain and asymmetrically experienced by caregivers providing more than 20 hours of weekly care. Further, our dynamic sequential matching results indicate that for caregivers providing multiple years of higher intensity care the negative effects persist.

1 Introduction

Ageing populations pose a serious challenge to health care systems in developed economies. The United Kingdom (UK) is an exemplary case, by 2050 more than a quarter of its population is expected to be 65+ and over 10% is predicted to be 80+ (OECD, 2019), drastically increasing the long-term care (LTC) demand (de la Maisonneuve & Martins, 2015). One solution to meet this demand is to rely on informal care, care provided by friends or relatives. Informal care is often preferred by the care-recipient and is from a governmental perspective a low-cost alternative to formal care. In addition, there is evidence that (partially) substituting formal by (unskilled) informal care does not jeopardize care-recipients' health. Receiving informal care can lower medical expenditures (van Houtven & Norton, 2004, 2008), decrease the likelihood of infections and bedsores (Coe et al., 2019) and improve recipients' mental health (Barnay & Juin, 2016). In the UK, informal care already plays a crucial role in meeting current care demand with more than 18% of the 50+ population providing informal care in contrast to the OECD average of 13.5% (OECD, 2019).

Despite these benefits there are concerns regarding the impact of informal care on caregivers' labor market and health outcomes. To make informed decisions on adapting current policies to future demands a thorough understanding of such effects is necessary. Previous studies either found no or negative effects of informal care provision on labor market outcomes (see Lilly et al. (2007) and Bauer and Sousa-Poza (2015) for reviews) but considerable health effects for the caregiver due to the mental and physical strain (see Bom et al., 2019a for a review).

Up to now most literature has focused on the immediate impact of care provision. However, it is also important to understand how these effects develop over time as many caregivers provide several years of care. According to the 2011 UK Census men and women at age 50 can expect to spend 4.9 and 5.9 years of their remaining life providing care (ONS, 2017). In light of this prospect, it is important to focus on health outcomes as conflicting hypotheses regarding the impact of duration of caregiving on health exist.[1] There are three opposing hypotheses regarding the association between the duration of informal care provision and the impact of care provision (see for overviews: Townsend et al. 1989, Haley & Pardo, 1989; Pinquart & Sorensen, 2003a). The first hypothesis is called the wear-and-tear hypothesis implying the impact of care provision will worsen over time as coping resources decline and care needs increase. For example, a prolonged exposure to stress arising from care tasks might deplete caregivers' resources to deal with the care strain. The trait hypothesis, on the other hand, suggests that the caregiving burden is constant. Even when health of the care-recipient deteriorates, caregivers maintain a constant level of adaptation. The care providers namely have pre-existing coping skills and

---

[1] A related literature focuses on the longer-term impact of caregiving on labor market outcomes. Schmitz & Westphal (2017) studied the German context and found informal care provision to have a longer-term impact on labor market outcomes, this effect did not differ dependent on the duration of care provision (e.g. individuals that provided 1 year of care compared to multiple years of care provision). Rellstab et al. (2020) studied the Dutch context and did not find any impact of care provision on labor market outcomes, which they argue might be attributable to the generous formal support system in the Netherlands.

resources which remain present during the care task. Lastly, the adaptation hypothesis argues that individuals learn to adapt to the situation. Following this theory, the negative impact of care provision will decline when care tasks are prolonged as caregivers develop new coping strategies or become less affected by the stress involved in their care tasks.

Some studies have already investigated how longer-term or high intensity informal caregiving is associated with health. In the UK for example, Hirst (2005), Legg et al. (2013), Vlachantoni et al. (2016) and Lacey et al. (2019) found a correlation between either long-term or high intensity care and negative health outcomes. However, these studies have difficulties in ascertaining the direction of causality: is poor health caused by the act of providing informal care or do individuals in poor health, e.g. due to old age, more often provide informal care? To study the causal impact of care provision on health over time one must account for endogeneity concerns resulting from the selection of individuals into informal caregiving. Bom and Stöckel (2021) used Dutch and UK panel data to explore whether health effects differ across long-term care systems of varying generosity in public care provision. However, they focused on the immediate effects of providing informal care while ignoring the temporal dimension of these effects and the dynamic nature of providing informal care over multiple years.

To our knowledge thus far only Schmitz and Westphal (2015) and De Zwart et al. (2017) have studied longer-term health effects of informal caregiving in a causal framework. Using German panel data and focusing on female caregivers, Schmitz and Westphal (2015) find negative mental health effects persisting for up to three years after care provision. De Zwart et al. (2017) used panel data from multiple continental European countries to explore the effect of spousal caretaking among the elderly population. They report negative effects on mental health and increased medical consumption in the first year after care provision. The disappearance of health effects over time could mean that caregiving effects do not last or that individuals find ways to cope with them, however, it might also result from selective attrition as individuals with demanding caregiving tasks are more likely to drop out of the panel surveys.

To better understand the longer-term health effects of care provision we explore the health effects of providing informal care in the UK context using data from the Understanding Society (USoc) longitudinal survey. We estimate both (i) the immediate and longer-term health effects of providing informal care for up to 5 years after the initial caregiving decision and (ii) the effect of providing additional years of care. These effects, and their relation to care intensity and caregiver characteristics, help policymakers to gauge the potential consequences of informal care provision and to identify those subgroups in largest need of support.

This study contributes to the literature on the longer-term health effects of providing informal care for the caregiver in several ways. The detailed individual-level information on caregivers and recipients allows us, unlike most previous studies focusing often on female or spousal caregivers, to explore the heterogeneity of caregiving effects across different groups of caregivers (e.g. by gender, care-recipient and intensity of care). Further, we estimate the health effects of multiple years of care provision using a dynamic matching approach (Lechner, 2009). The added benefit of this approach is that we can investigate the impact of additional years of care provision to determine how health effects evolve with continued caregiving. To our knowledge we provide the first causal estimates of caregiving effects in the UK that take the temporal dimension into account by exploring short- and longer-term health effects while explicitly addressing the dynamic nature of caregiving across multiple periods. While the UK is similar to Germany (studied by Schmitz & Westphal, 2015) with regards to the prominent role of informal caregivers in delivering social care services (Comas-Herrera et al., 2010), the countries differ in their generosity of caregiver allowances and formal care alternatives (Curry et al., 2019). Our results thereby also provide new evidence on the existence of caregiving effects and their magnitude from a different institutional context.

We find strong negative effects on mental health that are concentrated among high-intensity caregivers and remain persistent for multiple years. Additionally, the estimates from our dynamic matching procedure indicate that the mental health effect of care provision seems to persist for individuals providing care over multiple years. Using alternative outcome measures we confirm the consistency of our results and their economic relevance.

The paper proceeds as follows. Section 2 provides an overview of the UK long-term care system. The empirical strategy is outlined in Section 3, followed by a discussion of the underlying dataset in Section 4. Section 5 presents our results, starting with the baseline findings from a static model before proceeding to the dynamic matching approach. Robustness checks are described in Section 6, followed by a discussion and conclusion in Section 7.

2 Institutional Background

Formal LTC in the UK is organized in a mixed-system combining universal and means-tested benefits. Health services provided by the National Health Service (NHS) are free at the point of delivery and predominantly financed from taxation. The health-related components of LTC, which mostly entail nursing services, are funded via the NHS when granted by the GP (Comas-Herrera et al., 2010). Other types of LTC, such as residential care and help with personal tasks at home, are the responsibility of local authorities (Glendinning, 2013). Access to these services is dependent on locally determined needs-assessments. This care is offered via a safety-net structure requiring individuals to deplete their wealth before becoming eligible for publicly funded care (Colombo et al., 2011).[2] This system ensures that publicly funded LTC services are only provided to those with severe needs and unable to pay themselves (Fernández et al., 2009).[3] In 2015 the UK spent about 1.5% of its GDP on LTC with 23% of these expenditures related to social care (ONS, 2015).

As public LTC services are means-tested, a large part of LTC is provided informally with more than 18% of the UK 50+ population providing care (OECD, 2019). Additionally, more than a third of all caregivers do so for more than 20 hours per week according to data from the 2011 UK Census (ONS, 2013). In response to this large dependence on informal care, various policies to support informal caregivers (e.g. by providing information or support groups) are in place. The 2014 Care Act gave caregivers the right to receive a needs-assessment and corresponding support services (European Commission, 2018). However, reaching caregivers with this support is difficult as only six percent of caregivers receive any form of local authority support (Yeandle, 2016). Financial support is offered to informal caregivers via a "carer's allowance" (Carers UK, 2016). This allowance, amounting to £66.15 a week (approximately $86) in 2020 (UK Government, 2020) is paid to caregivers who meet restrictive conditions.[4] As the take-up of the allowance and its monetary value is low it is not a potent incentive to take up informal care for the related monetary gains (Colombo et al., 2011).

---

[2] Income and assets (including under certain circumstances housing wealth) are considered. Individuals with assets above GBP 23,250 are ineligible for support. Those with assets between GBP 14,250 - GBP 23,250 (approximately $18.448 - $30,100) are required to contribute to the costs while individuals with assets below GBP 14,250 have their costs completely covered (NHS, 2018).

[3] In case of self-funding expected costs are about £15/hour (approximately $19) for home care (Age UK, 2019a) and £600 and £800/week (approximately $777 and $1036) for care homes and nursing homes (Age UK, 2019b).

[4] Individuals can receive the carer's allowance when they (i) are aged 16 or over (ii) provide at least 35 hours of care a week; (iii) earn less than £123 per week (approximately $152); (iv) are not full-time students or studying for more than 21 hours a week; (v) normally live in the UK and have been in the UK for at least two of the last three years (UK Government, 2020).

3 Methods

The decision to provide informal care is not random. Individuals ‘select into’ informal caregiving, thereby creating endogeneity when studying its impact on health. We aim to overcome this problem by matching individuals on observable characteristics affecting health outcomes and the decision to provide informal care. To do so, we follow the intuition regarding the caregiving decision as proposed by Schmitz and Westphal (2015) who define three areas affecting the transition into informal care. The first are care obligations, as the most important determinant of informal care provision is the presence of a family member in need and the presence of alternative potential caregivers. The second category, willingness to provide care, refers to personality traits and socio-economic characteristics, as these affect individuals’ inclination towards providing care. Lastly, the ability to provide care refers to individuals’ own health status.

Our empirical strategy builds upon the potential outcomes framework by Rubin (1974) and addresses the endogeneity of providing informal care using regression adjusted propensity score matching (Rubin, 1979). The main assumption underlying propensity score matching is the conditional independence assumption (CIA). The CIA in our context states that after conditioning on a set of observable variables the potential health outcomes for both caregivers and non-caregivers are the same in the absence of informal care provision at all considered time periods. This implies that differences in health outcomes between caregivers and non-caregivers can be attributed to the provision of informal care. While it is only possible to match upon observed differences, this is not necessarily a problem as often unobserved characteristics are correlated to observed differences (Stuart, 2010). As we are able to match upon a broad range of variables related to informal care and health, we assume that we are able to capture unobserved covariates via their correlations with variables included in our propensity score regression. Additionally, following Lechner (2009), we exploit the panel structure of our data to match individuals upon information from the period directly preceding informal care provision to make this assumption more credible. The advantages of this strategy are that (i) providing care cannot affect the covariates and (ii) the previous caregiving status likely captures much of the unobserved heterogeneity affecting treatment assignment. For example, elements related to health status might be affected by past care provision while at the same time affecting current treatment assignment.

3.1 Static Matching

Our first aim is to estimate the longer-term impact of becoming an informal caregiver, abstracting from the question of the number of years someone provides care for. This static approach means that we match starting caregivers with non-caregivers and follow these two groups over time.

We identify individuals as treated when we observe their transition into caregiving, everyone who does not report any care-episode is included in the control group (untreated).[5]

Propensity scores of providing informal care are estimated using probit models. We estimate the propensity of providing low, medium or high intensity informal care conditional on the variables affecting the transition into care provision at t_{-1}. We use these propensity scores to match treated to untreated individuals. To increase the quality of the matching we estimate propensity scores separately by intensity group: all individuals providing care, irrespective of the reported intensity, and separately for the different intensity levels of low, medium and high intensity as defined in the data section. To make use of the large amount of information available in the dataset we use a kernel matching approach that uses weighted averaging on the untreated sample to form the counterfactual group.[6] We assess the common support, whether there is sufficient overlap in characteristics between the treated and untreated individuals, as the risk of kernel matching lies in the increased chance of including "bad matches", untreated individuals that are highly dissimilar to the treated group, in the estimation (Caliendo & Kopeining, 2008). Furthermore, as we do not match on actual covariates but on propensity scores, we assess whether balance of covariates is achieved after the matching procedure. We do so by using the standardized bias (Rosenbaum & Rubin, 1985). In our baseline specification we only match based on the estimated propensity scores, allowing us to include nearly all untreated individuals, therefore using more information and lowering variance (Caliendo & Kopeining, 2008). However, matching only on propensity scores itself has some drawbacks especially when a large range of control variables is used (Iacus et al., 2012). In the robustness checks section, we explore this issue in greater detail by changing our matching approach using precise one-to-one matching and coarsened exact matching (Blackwell et al. 2009)

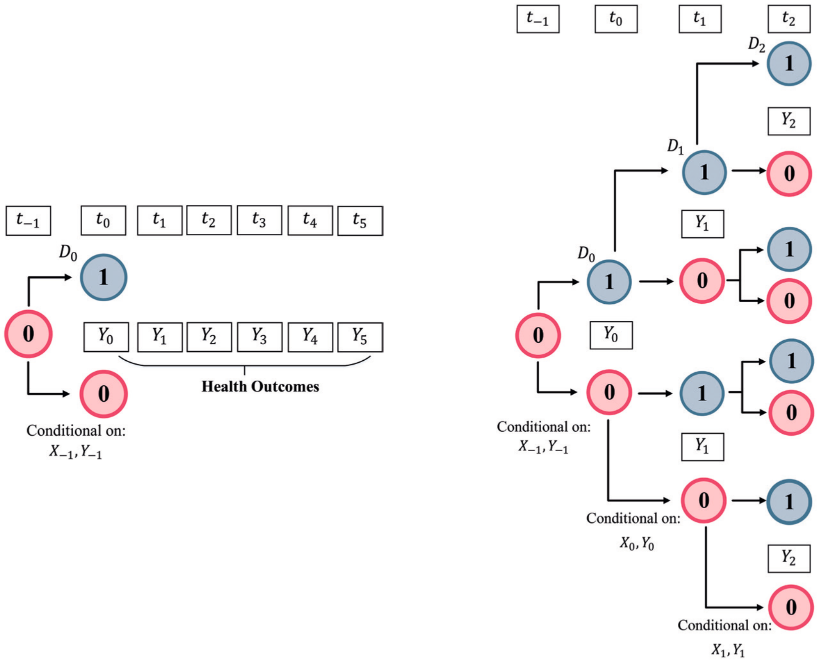
Finally, the average treatment effect on the treated (ATT) is estimated by regressing health outcomes on the treatment indicator (providing care) and all control variables used in the propensity score estimation with individuals in the control group weighted by their estimated kernel weights. By regressing on the control variables alongside the treatment indicator we aim to correct for remaining residual differences in the covariate distributions between the treatment and control group (Lechner, 2009; Rubin 1973). We do not use the covariates from later waves as these might be affected by the treatment. The health impact of providing care is estimated for the immediate time after first reported care provision and up to five years afterwards. Figure 1 provides a graphical representation of the static and dynamic matching designs.

---

[5] Future informal caregivers, although an ideal pool of suitable control group members, are not included in the control group due to the way our data is structured. To maximize the number of observable treated individuals we pool starting caregivers from across waves. To assess the robustness of this decision we also considered a situation where only caregivers starting to provide care within their first three years of survey participation are included, effectively moving half of the caregivers into the control group with t_{-1} being their respective entry wave. The results are depicted in Online Supplement Figure O6.A2.1 and are highly similar to our baseline results.

[6] We use the Stata command psmatch2 (Leuven & Sianesi, 2003) using an Epanechnikov kernel with a 0.03 bandwidth. The bandwidth choice is a trade-off between a small variance and an unbiased estimate of the true density function (Caliendo & Kopeining, 2008). While not reported in detail we have tested varying bandwidths, e.g. 0.01 and 0.06, with negligible impact on our results.

Figure 1: Matching Designs


a) Static Matching

b) Dynamic Matching

Source: Own illustration based on Schmitz & Westphal (2017). *Note:* D refers to the decision to either: (1) provide informal care or (0) not to provide informal care at a certain decision node. Y refers to the health outcomes, X refers to the included covariates.

3.2 Dynamic Sequential Matching

The static matching approach aims to answer the question *"If an individual starts to provide informal care in period* t_0 *(for an undefined time spell) does it change his or her health outcomes thereafter?"*. The treatment group hence contains individuals who stopped providing care in t_1 and those who continued caregiving for various years. This might bias the treatment effect estimates for periods past t_0 as they are partially based on individuals that no longer provide care. To explore to what extent the longer-term treatment effects are driven by multi-period caregivers we use a dynamic matching approach following the work of Lechner & Miquel (2010) and Schmitz & Westphal (2017).

In contrast to the static approach, the dynamic matching explicitly estimates the effect of providing a second (third) year of care while considering a potential effect of caregiving in t_0 (t_1) on health and other endogenous covariates that influence the decision to care provision in subsequent years, such as labour market status. It therefore helps to understand how the health effect of care provision is affected by duration of care and whether the static treatment effects over time are representative for the population of multi-period caregivers. Further, it allows us to answer the question whether caregivers adapt to their caregiving responsibilities over time.

The dynamic matching is computationally demanding. It requires the estimation of treatment probabilities at all possible decision nodes, thereby leading to 2^T possible treatment paths where T is the maximum possible treatment duration. Further, it requires the availability of all health outcomes and covariates at the time-period prior to (continued) caregiving as the matching is repeated at all decision nodes. We limit ourselves to the case of $T = 3$ (see Figure 1). This is motivated by two considerations. First, for our sample this time-window seems sufficient as the broad majority of caregivers provides a maximum of 3 years of consecutive care.[7] Second, in the dynamic framework time t is not defined relative to the first individual caregiving episode but fixed to allow for all potential treatment pathways to be observed, leading to less observable starting caregivers at t_0.

To illustrate the approach in more detail we provide an example showing the steps undertaken to estimate the marginal effect of providing two years of care instead of one. The treatment group in this example comprises everyone that provided informal care in both waves (t_0 and t_1), whereas the control group consist of everyone that provided care in the first wave (t_0) but not in the second (t_1). In the dynamic matching design in Figure 1 this refers to comparing the group that followed the path *0-1-1* with the group following the route *0-1-0*.

Consider a binary indicator D_t encoding care provision in period t. As in the static estimations, we start our analysis by estimating the propensity of providing informal care at the first node ($D_0 = 1$) conditional upon not providing care in the period before, and pre-treatment health outcomes and other covariates using a probit model. The propensity of providing informal care at the first node is: $Pr(D_0 = 1 \mid X_{-1}, Y_{-1})$. Therefore the resulting estimate is equivalent to the immediate effect (t_0) estimated in the static matching framework.

In extension we also estimate the decision taken at the second node ($D_1 = 1$) were caregivers decide to (dis-)continue caregiving. We estimate the propensity scores of both options conditional upon already being a caregiver and on health and the other observables both at the first and the second node. The propensity of providing informal care at the second node after providing care in the first period is: $Pr(D_1 = 1 \mid D_0 = 1, X_{-1}, Y_{-1,} X_0, Y_0)$. The propensity of discontinuing care provision is: $Pr(D_1 = 0 \mid D_0 = 1, X_{-1}, Y_{-1,} X_0, Y_0)$.

Just like in the static model, we estimate the propensity of providing informal care by care intensity, here medium and high intensity informal caregiving are combined due to sample restrictions. We use the propensity scores to calculate inverse probability weights (IPW). As IPW estimates might be sensitive to very high or low weights from individuals with very high or low propensity scores (Robins et al., 2000), we check whether our results are robust to

---

[7] Among the caregivers in our sample approximately 46.8% provide one year of care, 20.6% provide two and 13.0% provide three years of care. This leaves 19.6% of the caregivers providing more than three years of consecutive care. These numbers are based on caregivers starting in USoc waves 2-4 allowing all included respondents, in theory, to be able to have a caregiving spell of five or more years. Caregivers discontinuing their survey participation but continue to provide informal care are not observed.

removing observations with extreme weights as proposed by Lechner (2009).[8] Furthermore, for all scores we condition upon common support: in case no untreated counterparts with a similar propensity score for our treated respondents are present, these treated observations are excluded from the analysis.

Based on the estimated propensity scores we calculate inverse probability weights for both the treatment and the control group. These are defined as follows:

$$\frac{1}{\left(\Pr\left(D_0 = 1 | X_{-1}, Y_{-1}\right)\right) * \left(\Pr(D_1 = 1 \mid D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0)\right)} \; for\ the\ treatment\ group$$

$$\frac{1}{\left(\Pr\left(D_0 = 1 | X_{-1}, Y_{-1}\right)\right) * \left(\Pr(D_1 = 0 \mid D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0)\right)} \; for\ the\ control\ group$$

We estimate the dynamic average treatment effect on the treated (those who provide two years of care) by regressing health on the treatment while controlling for remaining differences by adding all covariates from the previous waves and weighting the data using the calculated inverse probability weights. We hence estimate, in this example, the health effects at t_1 of providing care in t_0 and t_1, compared to only providing care at t_0.

This sequential matching strategy was proposed by Lechner (2009) to estimate treatment effects in settings with dynamic treatment durations. While it follows a similar intuition as the static matching procedure identification is based on an augmented version of the CIA: the weak dynamic conditional independence assumption. Consider the case above when comparing outcomes of two and one years of informal care. The weak conditional independence assumption combines two parts. Firstly, the initial conditional independence assumption stating that potential outcomes in t_0 and t_1 are independent of treatment status in t_0 once we match upon observables at t_{-1}. Secondly, that potential outcomes in t_0 and t_1 are independent of continued treatment in t_1 once we condition on control variables and outcomes at both t_{-1} and t_0 and treatment status at the initial node t_0.

---

[8] We check our results using two approaches: (1) dropping scores for the first decision that are smaller than 5% or larger than 95% of the estimated propensity score distribution (2) dropping scores for the first decision that are smaller than 1% or larger than 99% of the estimated propensity score. Qualitatively our results are robust to these different specifications (results available upon request).

4 Data

We use data from the Understanding Society (USoc) dataset, also known as the UK Household Longitudinal Study (UKHLS; University of Essex, 2019); an annually conducted representative panel survey of the adult UK population (aged 16+). It started in 2009 with approximately 40,000 respondents across 30,000 households as the successor of the British Household Panel Survey (BHPS), which ended in 2008. In 2010, members of the last BHPS-wave were invited to join the USoc after which an additional 8,000 individuals joined. This paper uses all nine completed waves conducted between 2009 and 2019.[9]

Informal caregivers are identified using the question *"Is there anyone living with you who is sick, disabled or elderly whom you look after or give special help to (for example a sick, disabled or elderly relative/husband/wife/friend etc.)?"*. Individuals providing care outside their own household are identified based on the question *"Do you provide regular service or help for any sick, disabled or elderly person not living with you? [Exclude help provided in course of employment]"*. Apart from being able to identify individuals providing care inside and outside their own household the questionnaire also covers care intensity (hours per week) and the relationship between the caregiver and care-recipient.

We explore differences in the impact of caregiving dependent on the reported hours of care in a typical week as it is covered in the caregiving module of the annual USoc individual surveys. Based on these reported hours of care we split our sample of caregivers in three; low intensity (<10 hours of care per week), medium intensity (between 10 – 20 hours) and high intensity caregivers (more than 20 hours). When evaluating our results, one however must be aware of a potential downward bias in our estimates due to an underrepresentation of caregivers in the upper end of the intensity distribution. The share of high intensity caregivers in our sample (12.8%) is lower compared to the UK Census of 2011 which indicates that nation-wide about 37% of the caregivers provide care for more than 20 hours a week (ONS, 2013) or the 17% reported in the 2014 European Social Survey (ESS, 2014).

4.1 Health Outcomes

Various studies report the impact of care provision on mental and physical health (e.g. Pinquart & Sörensen, 2003b). To identify potential changes in both health domains we use the SF-12 health questionnaire in which individuals self-report on 12 questions related to various aspects of their own health in the past four weeks. From the survey we derive the physical (PCS) and mental (MCS) component summary scores which are constructed using different subscales related to physical and mental health.[10] The two health scales are validated for the UK context

---

[9] It follows an overlapping panel structure where waves cover two years but overlap for one and individual respondents are surveyed every 12 months. Therefore, nine waves are available for the 10-year time period.

[10] The PCS comprises the subscales: Physical functioning, Role-Physical, Bodily Pain and General Health. The MCS comprises the subscales: Vitality, Social Functioning, Role-Emotional and Mental Health.

and range from 0 to 100, where a higher score represents better health. By construction MCS and PCS scores have a mean of 50 and standard deviation of 10 (Ware et al., 1995).

4.2 Time Structure

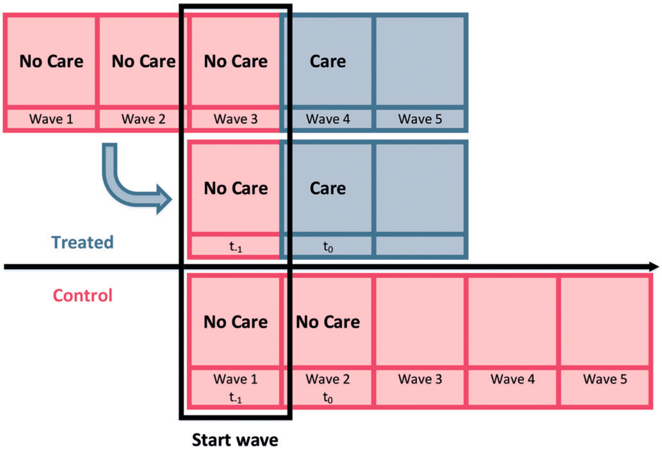
For the static matching procedure, we define a relative time variable depending on an individual's first reported care-episode as observed in the sample, meaning the first time an individual reports to provide care as part of his or her individual observation period as a survey participant.[11] Figure 2 provides a visualization of this time structure for the case of an individual entering the survey in wave 1 reporting their first care episode in wave 4. Among caregivers t_{-1} is defined as the period before the first reported caregiving episode. For everyone in the control group t_{-1} is the individual's first appearance as a survey participant in absence of any care episode during their participation. This time structure is chosen to maximize the number of observable treated individuals.

The analysis sample for the dynamic specification uses an augmented time-structure to allow for the modeling of all decision nodes between t_{-1} and t_2 and the comparison of various care trajectories. The time variable is normalized to t_{-1} being the entry wave of an individual into the panel for those who provide no care at any time-point and caregivers who start providing care within the first four participation waves. To increase the number of observable caregivers in the different caregiving trajectories we additionally include individuals whose caregiving trajectory starts after at least four periods of not providing informal care. For these the fifth participation is defined as t_{-1}. We therefore emulate the time-structuring in the static design by pooling caregivers from different starting waves but allowing these to have caregiving spells of up to three years. The important difference is that not all care-giving trajectories start at t_0 and at all decision nodes the control group contains individuals who will transition into a caregiving spell in future periods.

---

[11] Ideally, we would like to ensure that individuals observed as becoming first-time caregivers did not provide informal care previous to their participation in USoc which by definition we cannot observe. Unobserved previous caregiving status and its potential effect on certain covariates such as labour market participation or health outcomes would result in a violation of the conditional independence assumption. This cannot be checked without explicitly asking respondents about previous caregiving. As also among the control group individuals might have provided care before participating in USoc we do not think that this leads to a relevant bias in practice.

Figure 2: Static Dataset – Time structure example



Source: Own illustration.

4.3 Sample Selection

We construct two distinct datasets to implement the static and dynamic matching procedures. Individuals who identified as caregivers in their first observation period are excluded as transition into caregiving is not observed. For the static estimation we include all individuals that provide information for at least two time points on their health outcomes (t_{-1} and t_0) and provide full information on all covariates used in the propensity score estimation at t_{-1}. Individuals here predominantly drop out of the sample because they are proxy respondents or fail to provide sufficient information on their health outcomes or existing family members.[12] Individuals remain in the sample during the subsequent time points t_1 to t_5 in case information on the outcome-variables is available. For the dynamic estimation procedure, data requirements are more restrictive as we re-estimate propensity scores at each decision node. For this analysis, only individuals with complete information on all control variables for three waves (t_{-1} to t_1) and full information on outcome variables for four consecutive periods are included in the sample. Table O6.A1.1 in the Online Supplement provides an overview of the exclusion criteria applied to the analysis samples.

---

[12] Co-habiting family members are observed at every wave but family networks outside of respondents' households are only inquired about every second wave. If living family members are reported in a future wave this is applied backwards where information is missing. Likewise, if a family member is reported as deceased this is applied forward.

5 Results

5.1 Matching Quality

The descriptive statistics for the static matching sample are depicted in Table 1. Before the propensity score matching there is strong covariate imbalance between the control and treatment groups. For example, the individuals in the treatment group are on average almost nine years older and in worse health than their counterparts in the control groups across all health measures. We calculate the standardized bias for each covariate by taking the difference in means between the treatment and control group and dividing it by the square root of the average sample variance control and treatment groups (Rosenbaum & Rubin, 1985). This provides a standardized measure of the difference between both groups expressed as percentage points[13]. We follow the rule of thumb suggested by Caliendo & Kopeinig (2008) which states that there is sufficient balance when the bias is below 3-5%. Prior to the matching there is considerable imbalance between caregivers and the control pool. Matching corrects this imbalance, as can be seen in the post-matching differences between the groups. Details on the matching can be found in Online Supplement O6.A3. Online Supplement Table O6.A3.1 compares the pre- and post-treatment covariate balance and provides evidence for the large imbalance between treatment and control groups but also within the treatment group across the different levels of caregiving intensity. The matching by intensity group balances the covariates in all groups. Online Supplement Figure O6.A3.1 depicts for each intensity group the overlap in propensity scores as well as the imbalance pre- and post-matching graphically, providing evidence that the matching is successful. In each specification a small number of individuals is identified as off-support and dropped from the analysis: four in the any-care specification, none for low-intensity caregivers only in the treatment group and three and four for medium- and high-intensity caregivers. The corresponding information for the dynamic matching is displayed in Online Supplement O6.A4.

Table 1: Descriptive Statistics and Matching Quality

	Summary Statistics					Post-Matching Differences			
	Control Pool		**Treated**			**Bias in %**			
	Mean	SD	Mean	SD	**% Bias**	Any Care	<10h	10-20h	>20h
Mother alive	0.70	(0.46)	0.59	(0.49)	**-22.00**	0.40	0.60	-2.80	-3.40
Mother age	59.68	(11.28)	68.12	(12.71)	-4.00	0.80	1.10	-1.80	-2.70
Father alive	0.63	(0.48)	0.45	(0.50)	**-36.50**	1.30	1.40	-1.00	-2.90
Father age	60.99	(10.76)	68.01	(12.08)	**-23.70**	1.90	2.30	0.60	-2.30
Both parents alive	0.59	(0.49)	0.39	(0.49)	**-40.90**	1.00	1.30	-2.10	-2.70
Siblings alive	0.88	(0.32)	0.87	(0.34)	**-5.60**	0.10	0.60	-0.10	-3.60
Partner existing	0.63	(0.48)	0.70	(0.46)	**13.90**	0.50	0.30	0.20	1.70

---

[13] The exact formula is $(\bar{x}_{D=1} - \bar{x}_{D=0})/\sqrt{0.5 * (V(x_{D=1}) + V(x_{D=0}))}$ where x is the control variable of interest, V (x) is the sample variance for the treated (D = 1) and control (D = 0) groups.

Partner age	43.67	(14.13)	52.75	(14.61)	**35.90**	1.00	0.10	2.00	4.50
Own age	41.37	(17.14)	50.53	(16.14)	**55.00**	0.10	-0.10	4.00	5.20
Female	0.53	(0.50)	0.59	(0.49)	**13.10**	0.00	0.00	2.20	0.40
Children in HH	0.39	(0.49)	0.31	(0.46)	**-16.60**	-1.10	-0.40	-1.70	0.30
Children <14 in HH	0.35	(0.48)	0.28	(0.45)	**-16.40**	-0.90	-0.20	-1.40	0.20
Education: Primary	0.18	(0.38)	0.21	(0.41)	**9.20**	-1.60	-1.50	-1.10	-0.30
Education: Secondary	0.44	(0.50)	0.43	(0.50)	-1.40	-0.30	-0.70	1.80	0.80
Education: Tertiary	0.39	(0.49)	0.36	(0.48)	**-6.10**	1.70	2.00	-0.90	-0.50
Any paid work	0.63	(0.48)	0.57	(0.50)	**-14.00**	2.80	2.80	0.90	0.20
Full time work	0.50	(0.50)	0.40	(0.49)	**-21.10**	2.90	2.90	-0.70	-1.60
Self employed	0.07	(0.25)	0.08	(0.27)	3.90	0.20	0.60	0.00	0.30
Unemployed	0.05	(0.22)	0.04	(0.21)	-2.70	-1.20	-0.90	0.70	0.40
Homecarer	0.05	(0.22)	0.06	(0.24)	**5.10**	-2.30	-1.40	-1.40	-1.70
Disabled	0.02	(0.15)	0.04	(0.19)	**8.20**	-1.90	-2.20	-2.90	-2.70
Retired	0.14	(0.35)	0.26	(0.44)	**29.30**	-0.70	-1.20	2.30	3.70
HH income (1000€)	1.65	(1.17)	1.69	(1.22)	3.30	3.30	2.80	1.60	1.00
Big 5: Openness	4.59	(1.27)	4.58	(1.32)	-1.10	1.10	1.10	-1.10	-0.10
Big 5: Conscientiousness	5.45	(1.09)	5.57	(1.10)	**11.20**	2.20	1.90	2.00	1.40
Big 5: Extroversion	4.59	(1.30)	4.62	(1.29)	2.80	-0.70	-0.50	-1.30	-1.50
Big 5: Agreeableness	5.57	(1.03)	5.69	(1.02)	**11.60**	1.00	0.60	2.20	1.30
Big 5: Neuroticism	3.56	(1.43)	3.53	(1.45)	-1.90	0.40	-0.30	-0.20	1.20
Self Assessed Health	2.42	(1.06)	2.58	(1.05)	**15.10**	-2.80	-2.50	-1.50	-2.40
Mental Health Score	51.09	(9.10)	49.98	(9.85)	**-11.80**	1.70	1.80	0.70	1.80
Physical Health Score	51.56	(9.97)	49.46	(10.86)	**-20.10**	3.00	2.70	1.80	2.20
Long Standing Illness	0.29	(0.45)	0.37	(0.48)	**18.00**	-2.20	-2.10	0.10	-1.60
Number of Limitations	0.44	(1.22)	0.55	(1.29)	**8.50**	-3.00	-2.60	-1.70	-3.20
Satisfaction with Health	5.09	(1.60)	4.73	(1.72)	**-21.60**	2.10	1.60	0.60	1.00
Satisfaction with Income	4.64	(1.64)	4.51	(1.69)	**-7.80**	2.90	2.00	1.50	1.20
Satisfaction with Time	4.77	(1.62)	4.79	(1.68)	0.90	0.50	-0.10	0.80	0.50
Satisfaction with Life	5.37	(1.37)	5.20	(1.47)	**-11.60**	2.00	1.70	1.00	-0.10
Inversed GHQ Score	10.41	(2.73)	10.26	(3.01)	**-5.20**	0.80	0.90	1.00	-0.30
Individuals		**13,141**		**7,106**		**7,102**	**5,249**	**812**	**934**

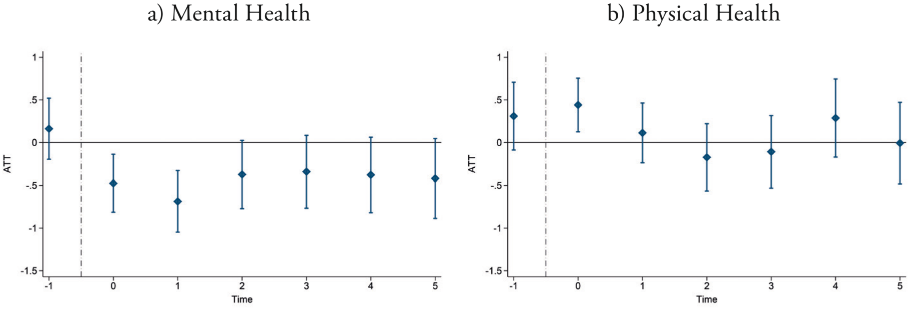
Source: USoc Waves 1-9, own calculations. *Note:* The number of informal caregivers (treated) in the different intensity groups do not add up to 7,106 as 104 caregivers provide unclear information regarding the amount of care provided. Individuals off-support are not included in the post-matching groups. Regional dummies are omitted. Big 5 variables range from 1-7 (low-high); SAH ranges from 1-5 (excellent-poor); satisfaction questions range from 1-7 (completely dissatisfied-completely satisfied); SF-12 Mental and Physical Component Scores range from 0 to 100 (lowest – highest level of health); inversed General Health Questionnaire (GHQ) score ranges from 0-12 (most distressed - least distressed).

5.2 Static Matching Results – Treatment Effects by Care Intensity

The results of the baseline static matching procedure by intensity will be presented graphically. All underlying estimates are reported in Online Supplement Table O6.A1.2. While the graphs depict the full sample results by caregiving intensity, we will in text also discuss the results by gender which can be found in Online Supplement Table O6.A1.2. In the baseline analysis we estimate the effect of any informal care provision irrespective of the reported intensity. Figure 3 depicts the estimated ATTs on both the (a) mental and (b) physical health scores across time and their 95% confidence intervals. To show that matching resolved any pre-treatment differences in health between both groups, the graphs depict the pre-treatment estimate at t_0. As can be seen in all graphs, before commencing of care provision no differences in physical or mental health are present between the matched groups.

To ease interpreting estimated coefficients with respect the magnitude of reported ATT estimates we report the corresponding effect size in percentage points of the standard deviation in brackets.[14] In the first years after the start of care provision, we observe small immediate negative effects in the mental domain of -0.476 ($p<0.01$; 5.07% SD) at t_0 and -0.687 ($p<0.001$; 7.32% SD) at t_1. In later periods these effects remain negative although not significant at the 5% level. In the physical domain baseline estimates indicate a small positive effect of 0.441 ($p<0.01$; 4.26% SD) at t_0 but no effects thereafter. The separate analyses by gender show that the results are driven by female caregivers. Female caregivers experience small negative mental health effects of -0.507 ($p<0.01$; 5.17% SD) at t_0 and -0.915 ($p<0.01$; 9.34% SD) at t_1 and -0.595 ($p<0.05$; 6.07% SD) at t_3. Male caregivers, do not experience consistent mental health effects but we observe a small positive impact on the PCS of 0.507 ($p<0.05$; 5.14% SD) at t_0 and 0.870 ($p<0.05$; 8.83% SD) and 0.764 ($p<0.05$; 7.76% SD) at t_4 and t_5.

Figure 3: Health Effects of Any Care Provision


a) Mental Health
b) Physical Health

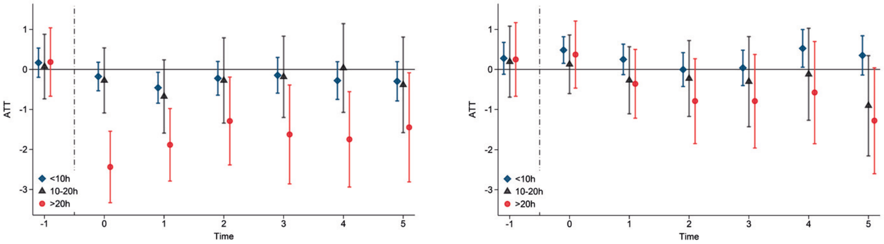
Source: USoc Waves 1-9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

---

[14] We use the unweighted pre-treatment standard deviation estimates for the mental and physical component scores for the respective analysis samples.

To explore heterogeneities in the estimated treatment effects we subdivide caregivers into treatment groups according to the reported weekly hours of care. Figure 4 plots the results by care intensity for (a) mental and (b) physical health. Low intensity caregiving for less than 10 hours per week is depicted in blue, medium intensity between 10 and 20 hours per week in grey, and high intensity caregiving of 20 hours and more in red. Low intensity care is the most commonly observed with 5,249 individuals (74%), followed by 934 high (13%) and 812 (11%) medium intensity caregivers.[15] By construction, effects in later periods are less precisely estimated as not all caregivers are observed for all years. At t_5 only 1,881 or 36% of the initial low intensity caregivers are still observable, for medium intensity this is 296 (36%) but only 293 (31%) high-intensity caregivers.

Figure 4: Health Effects by Intensity of Care Provision



a) Mental Health

b) Physical Health

Source: USoc Waves 1-9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

The large heterogeneity in the estimated treatment effects underline the importance of care intensity. In the mental domain we do not find effects for low or medium intensity caregivers, although all estimates point towards the negative direction. Among individuals providing care for more than 20 hours per week we observe strong initial negative effects of -2.438 ($p<0.001$; 26.24% SD) at t_0 and -1.884 ($p<0.001$; 20.27% SD) at t_1. While these effects decrease for subsequent periods, they remain largely persistent with -1.290 ($p<0.05$; 13.86% SD), -1.625 ($p<0.05$; 17.49% SD), -1.748 ($p<0.01$; 18.82% SD) and -1.448 ($p<0.01$; 15.59% SD) at t_2 to t_5. The previously apparent differences in caregiving effect by gender decrease when stratifying the samples by care-intensity. Both male and female high-intensity caregivers experience comparable negative mental health effects following transition into caregiving; -2.471 ($p<0.001$; 25.40% SD) for females and -2.396 ($p<0.001$; 27.66% SD) for males at t_0.

In the physical domain the pattern across care intensity levels is different. For low intensity caregivers we find a small positive immediate effect of 0.441 ($p<0.01$; 4.36% SD) at t_0 while for the other intensity groups the coefficient is similar but insignificant. At subsequent periods the

---

[15] Caregivers providing no intensity information are excluded (104). Detailed sample sizes by time-period t can be found in Online Supplement Table O6.A1.2.

estimated effects vary considerably. For low intensity caregivers the estimated coefficients are insignificant while varying around zero, except for t_4 in which we observe a positive effect of 0.561 ($p<0.05$; 5.55% SD). For medium and high intensity caregivers' coefficients are negative throughout except for the positive estimates at t_0 but not significantly different from zero. The coefficients for high intensity care seem to follow a downward trend. In the Online Supplement we provide evidence that this pattern seems driven by age-dependent physical health trends captured inadequately in the propensity score based matching (see Robustness Checks and Online Supplement Figures O6.A2.2 and O6.A2.3).[16]

5.3 Static Matching Results – Treatment Effects by Caregiver-Recipient Relationship

While the results in Figures 3 and 4 suggest that the negative effects of providing informal care are mediated by the intensity of care provided an alternative mediating factor could also be the relationship between caregiver and recipient (Bobinac et al., 2010). To explore this, we include an interaction term for the caregiver care-recipient relationship. In a first step we do so for the specification in which all caregivers are in the treatment group, irrespective of the reported intensity. As shown in the upper half of Table 2 in this specification it seems that while informal care itself has not effects on either mental or physical health the provision of care to one's partner does have a strong, negative effect on mental health. The interaction term for parental care on the other hand is small and insignificant across all time-periods.

However, when doing the same exercise but only on the group of caregivers for which our baseline specification finds consistent and substantial negative mental health effects, those providing high-intensity care, this is not observed. These results are shown in the lower half of Table 2. The reason for this pattern is likely found in the different distribution of intensity by informal caregiver-recipient relationship. While care to parents makes up 41% of all caregivers it is only 22% among high intensity caregivers. Care to a partner, which almost exclusively happens within the own household, makes up only 15% of the overall population of caregivers but 40% of the high-intensity caregiving sample.

---

[16] As propensity scores are a summary measure estimated using many covariates, these age-related trends are not guaranteed to be perfectly captured, e.g. when including age continuously instead of age-group dummies that capture different trends across age-groups. For example, a younger individual might receive a high propensity score due to his/her physical health being low and/or other strong predictors but would be faced with an entirely different physical health trajectory in the short and medium term compared to an older individual. As illustrated in the Online Supplement this age-dependent trend is not present for mental health.

Table 2: Treatment Effects Including Care-Relationship Interaction Terms

	Any Care											
	t_0		t_1		t_2		t_3		t_4		t_5	
	MCS	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**
Any Care	-0.207	0.655	-0.458	0.196	-0.294	-0.122	0.043	-0.133	-0.135	0.447	-0.385	0.152
	(0.201)	(0.183)	(0.221)	(0.212)	(0.246)	(0.236)	(0.258)	(0.262)	(0.279)	(0.289)	(0.306)	(0.314)
Spousal Care	**-1.640*****	**-0.754****	**-1.543*****	**-0.766***	**-1.193****	-0.561	**-1.679****	-0.502	**-1.712****	-0.573	**-1.508***	-0.888
	(0.301)	(0.291)	(0.346)	(0.324)	(0.379)	(0.389)	(0.451)	(0.445)	(0.501)	(0.521)	(0.617)	(0.583)
Parental Care	-0.082	-0.249	-0.025	0.060	0.198	0.062	-0.355	0.213	-0.032	-0.187	0.374	-0.089
	(0.216)	(0.194)	(0.246)	(0.230)	(0.273)	(0.253)	(0.300)	(0.287)	(0.342)	(0.329)	(0.383)	(0.376)
Control	**1,341**		**12,497**		**10,575**		**9,753**		**8,701**		**8,185**	
Treatment	**7,102**		**5,863**		**4,901**		**4,200**		**3,239**		**2,503**	
Spousal Care	1,042		830		677		546		429		322	
Parental Care	2,937		2,477		2,088		1,806		1,409		1,112	
	High Intensity Care											
	t_0		t_1		t_2		t_3		t_4		t_5	
	MCS	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**
Any Care	**-2.486*****	0.133	**-1.724***	-0.635	-1.038	-1.100	-1.051	-1.531	-1.430	-0.765	-1.714	-1.872
	(0.604)	(0.537)	(0.676)	(0.603)	(0.763)	(0.673)	(0.831)	(0.787)	(0.846)	(0.871)	(0.975)	(0.942)
Spousal Care	-0.056	0.224	-0.509	0.108	-0.888	-0.082	-1.273	0.944	-1.143	0.370	-1.153	1.065
	(0.689)	(0.617)	(0.752)	(0.702)	(0.809)	(0.790)	(1.003)	(0.905)	(1.083)	(1.130)	(1.408)	(1.249)
Parental Care	0.291	0.623	0.178	0.992	0.404	1.428	-0.349	1.612	0.534	0.189	3.095	0.817
	(0.836)	(0.687)	(0.903)	(0.801)	(1.022)	(0.876)	(1.190)	(1.007)	(1.225)	(1.197)	(1.513)	(1.379)
Control	**13,141**		**12,497**		**10,575**		**9,753**		**8,701**		**8,185**	
Treatment	934		722		606		488			384	293	
Spousal Care	374		286		239		185		149		110	
Parental Care	207		167		149		115		94		68	

Source: USoc Waves 1-9, own calculations. *Note:* *p<0.05, ** p<0.01, ***p<0.001, standard errors in parentheses. MCS columns depict the estimated coefficients for the SF12 Mental Component Scores, PCS the SF12 Physical Component Scores.

5.4 Dynamic Matching Results – Treatment Effects by Care Intensity

Next to estimating the impact of at least one year of care provision, we aim to investigate the impact of providing additional years of informal care and the extent to which the static ATT estimates are driven by ignored changes in treatment status over time. For this dynamic matching approach, we estimate the propensity of (not) providing informal care at every decision node and drop scores in case the observation is off support. In Online Supplement O6.A4 we report the propensity scores for the different care-trajectories as well as an overview of the excluded individuals. Further, we estimate the treatment effects using the static matching approach for the same sample used in the dynamic matching to provide a comparison between both estimation strategies. For the dynamic matching we merged the groups of medium and high intensity caregivers due to concerns about statistical power. Further we excluded individuals with unstable care trajectories to not wrongfully capture the impact of increasing care intensity among continuing caregivers. The results for both mental and physical health are depicted in Table 2. Due to data limitations, we are unable to estimate the propensity of providing medium or high intensity care after the second node. Most individuals continue care provision when already done so for two years in a row and with a higher intensity. This limited variation in care-continuation among this groups obstructs us from running these models which are highly data-demanding as they include covariates from all previous waves. In the below presented results we hence only present the estimates for any and low-intensity caregivers at the 3rd year.

The first column of Table 3 depicts the estimated caregiving effect at t_0 for the first period of informal care. For mental health both dynamic and static matching indicate a small and insignificant negative coefficient for any care provided of -0.312. When separating the different intensity levels there are no significant differences for low intensity caregivers while for medium or high intensity caregivers the estimates are negative and significant with -1.374 ($p<0.01$; 14.66% SD). At t_1 the static matching, which pools both continuing and discontinuing caregivers together, again indicates a continuing negative mental health effect of -1.413 ($p<0.001$; 15.08% SD). The dynamic matching estimates, which account for the impact of previous-period caregiving on covariates and health outcomes, indicate that the health effects among continuing caregivers are larger. When focusing on the group of continuing caregivers the difference is -2.254 ($p<0.001$; 24.05% SD) when using the never caregivers as a control group. Explicitly comparing one against two years of medium/high intensity care results in an estimated health impact of the second year of care provision that is again larger at -1.836 ($p<0.05$; 19.59% SD) although less precisely estimated.

Table 3: Dynamic Matching Estimates

			Mental Health				
	t_0	t_1			t_2		
	Static/ Dynamic	**Static**	**Dynamic (2v0)**	**Dynamic (2v1)**	**Static**	**Dynamic (3v0)**	**Dynamic (3v2)**
Any	-0.312	-0.394*	-0.350	-0.151	-0.190	0.0195	0.610
	(0.194)	(0.201)	(0.315)	(0.405)	(0.202)	(0.405)	(0.771)
Low Intensity	-0.007	-0.116	-0.103	0.197	0.090	-0.215	0.582
	(0.209)	(0.218)	(0.323)	(0.451)	(0.224)	(0.518)	(0.829)
Medium/ High Intensity	**-1.374****	**-1.413****	**-2.254*****	**-1.836***	**-1.148***		
	(0.462)	**(0.470)**	**(0.546)**	**(0.810)**	**(0.452)**		
			Physical Health				
	t_0	t_1			t_2		
	Static/ Dynamic	**Static**	**Dynamic (2v0)**	**Dynamic (2v1)**	**Static**	**Dynamic (3v0)**	**Dynamic (3v2)**
Any	0.178	-0.148	0.012	0.144	0.059	0.014	-0.147
	(0.173)	(0.191)	(0.283)	(0.383)	(0.200)	(0.361)	(0.574)
Low Intensity	0.340	0.016	0.080	0.094	0.355	0.477	-0.073
	(0.193)	(0.217)	(0.303)	(0.437)	(0.224)	(0.375)	(0.622)
Medium/ High Intensity	-0.334	-0.707	**1.379****	1.267	**-1.010****		
	(0.387)	(0.390)	**(0.506)**	(0.800)	**(0.424)**		
Treatment (Control)	1672 (18812)	1672 (18812)	700 (17268)	700 (967)	1672 (18812)	406 (15756)	406 (274)
Low Intensity	1285 (18812)	1285 (18812)	536 (17260)	536 (738)	1285 (18812)	296 (15865)	296 (223)
Medium/ High Intensity	313 (16081)	313 (16081)	146 (16660)	146 (186)	313 (16081)		

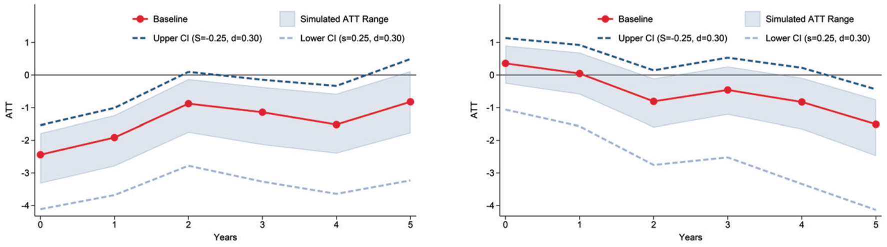
Source: USoc, own calculations. *Note:* The static and dynamic results at t=0 slightly differ at the second decimal as either matching or inverse probability weighting is used, the depicted results are the static results. The table presents the dynamic ATT, the effect of providing an additional year of care compared to an individual following an alternative care trajectory. A second/third year of informal care is compared to: not providing informal care (2v0 or 3v0); providing care for only one year (2v1); or providing care for two years (3v2). It compares the health of treated and matched controls based on the information from the directly preceding wave. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors are in parentheses. Please note that the displayed static results are based on a different sample than the baseline static results as a differently conditioned sample is used. See the Appendix for more details.

For physical health no effects are found at t_0, nor among the any and low estimates at any of the other timepoints. Later on, we however observe mixed results. Using the static approach, a negative impact at t_2 is observed. Focusing on the dynamic results, the estimates at the second node however point into the positive direction, and turn significant when comparing individuals observed for two years of care compared to their non-caregiving counterparts. While this might indicate a positive impact of continuing care provision, we would like to be cautious in interpreting these results in particular. Among this small group matching at all nodes becomes difficult. This seems specifically the case among results in the physical health dimension for high intensity caregivers. As presented in the robustness checks section these effects are most sensitive to alternative model specifications. Additionally, please note for all of these estimates, the standard errors rather large. This makes it for many of these estimates, impossible to differentiate between a zero-effect or a non-significant effect.

6 Robustness Checks

There are two primary concerns related to the estimation of causal effects using matching-based estimators. Firstly, a violation of the main identifying assumption, the conditional independence assumption (CIA), and secondly the model dependence of estimated treatment effects to the specific way in which matches are obtained. The CIA is an inherently untestable assumption and hence one has to rely on data-driven methods to explore the extent to which the observed treatment effects could be reasonably explained by an omitted variable that effects both selection into treatment and (health) outcomes in absence of treatment. We applied the simulation-based method proposed by Ichino et al. (2008) to explore the sensitivity of our results to such an omitted variable having both a selection effect (s) and an outcome effect (d). An example for such an omitted variable could be the latent health of a family member leading to both an increased likelihood to provide care (selection effect) as well as lower mental health irrespective of treatment uptake (outcome). In Online Supplement Section O6.A2 we provide a detailed description of the procedure. In short, the simulation re-estimates the propensity scores leaving out one of our covariates to estimate the space of selection and outcome effects among our pool of covariates to obtain a range for s and d that is plausible (see Online Supplement Table O6.A2.1). From this range we construct a confounder that has a combination of selection and outcome effects that would bias our estimated ATTs severely into either direction by choosing a combination of the highest observed *s* and *d* among our included control variables. Based on this we then calculate a range of ATTs and corresponding confidence intervals to obtain a bound on the treatment effects and their sensitivity to omitted variables. Figure 5 depicts both of these for mental and physical health for the high-intensity caregiving group.

Figure 5: Simulated Violation of the CIA


a) Mental Health

b) Physical Health

Source: USoc Waves 1-9, own calculations.

While there is clear evidence for a simulated confounder to shift the overall range of estimated treatment effects especially the mental health effects for the initial periods after providing informal care decrease in size but remain strongly significant.

To address the second primary concern with respect to the sensitivity of our results to our matching procedure we also explore different ways to obtain matched treatment and control groups. As discussed by Iacus et al. (2012) one primary concern with propensity score based methods is the tendency that high dimensionality in the chosen covariates in can result in increasing imbalance in covariate values, even if the one-dimensional propensity score differences are minimal. We address this by using coarsened exact matching (Blackwell et al., 2009) to pre-process the data. Coarsened exact matching pre-processes the treatment and control group by creating strata from both samples which are highly similar based on a selection of covariates that should be balanced under all circumstances before conducting the propensity score estimation to weigh observations within these strata.

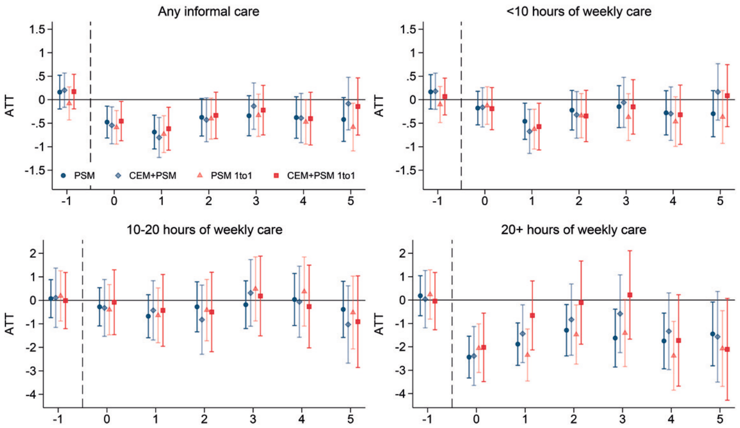
The covariates we choose for the coarsened exact matching are the age and living status of the parents, gender, the presence of a long-standing illness or disability, the respondents age and lastly mental and physical component scores at t_{-1}. This reduces the number of included control and treatment individuals considerably as they are not within the strata in which these covariates would be balanced.[17] After reducing the sample, the regular propensity score matching is performed in the same way as in our baseline specification. Next to this coarsened exact matching with subsequent propensity score matching we additional perform a one-to-one matching using propensity scores only using the reduced sample after the coarsened exact matching and the baseline specification. Figure 6 depicts for both mental and physical health the ATT estimates for all specifications and by treatment intensity. The matching quality for each is similar across specifications while details can be found in Online Supplement O6.A3 Figures O6.A3.2 to O6.A3.4. The results depicted in Figure 6 suggests that across specifications the mental health effects are largely similar especially for the initial periods after care-provision has started. However, with regards to the longer-term effects the matching procedure seems to have a stronger influence on the resulting ATT estimates.

---

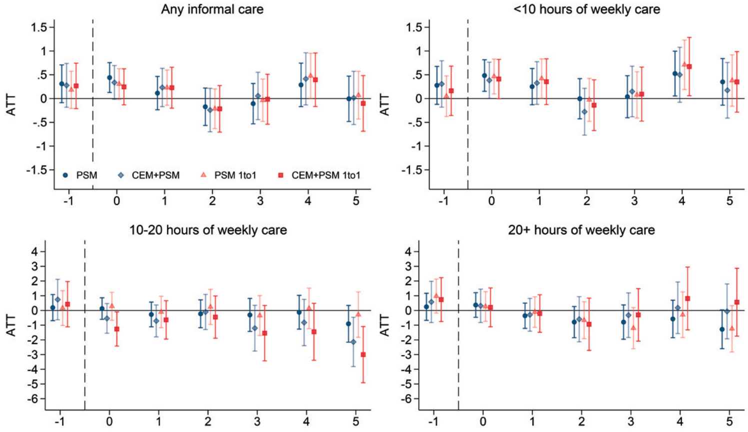
[17] The remaining number of individuals in the groups by treatment intensity are for control/treated: 7,549/4,162 for any care, 6,797/3,149 for low intensity, 2,205/446 for medium intensity, and 2,364/534 for high intensity.

Figure 6: Results by Matching Procedure

a) Mental Health Effects by Matching Procedure



b) Physical Health Effects by Matching Procedure



Source: USoc Waves 1-9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

Besides these two main concerns we conduct a range of additional analysis to address other related concerns about our identification strategy. For the sake of brevity, we only shortly discuss these with detailed results available in the Online Supplement. The observed mental health effects could be explained by an ongoing trend that started before informal care provision. Due to our time-structure we observe for many caregivers their mental and physical health for multiple periods prior to providing care. Online Supplement Figure O6.A2.3 plots the mean MCS and PCS for all intensity levels for up to four years prior to providing informal care. There is little evidence that the observed results are driven by a distinct negative trend in caregivers' mental health before the actual onset of care provision.

Further, we explore the existence of a potential downward bias underlying our results due to selective attrition. We follow De Zwart et al. (2017) by splitting our sample into two groups and re-estimating the initial treatment effects. For the first group we observe health states past t_1, while for the second group we can only observe health immediately after providing informal care due to permanent survey attrition. Online Supplement Figure O6.A2.4 plots the treatment effects for both groups and indicates that the attrition sample experiences more persistent negative effects directly before discontinuing their participation, but no physical health effects. These results indicate some evidence for a downward bias in our estimated treatment effects for mental health in later periods.

While the SF-12 component scores allow us to measure mental and physical health, the interpretation of effect sizes is not straightforward. To do so we use an alternative outcome measure, the general health questionnaire (GHQ), which is a mental health screening instrument with defined thresholds identifying individuals at risk of developing a mental illness. Online Supplement Figure O6.A2.5 present the corresponding results. When considering this alternative outcome measures our results remain generally the same, indicating an asymmetric effect on mental health especially among high-intensity caregivers. However, the results for GHQ scores depict a pronounced dose-response relationship, not observed when using MCS as the mental health measure. Further, these results also indicate that the decrease in mental health is economically relevant as the number of individuals with surpassing screening thresholds increases substantially by 4 to 8% depending on the used measure and applied threshold.

A concern for our dynamic matching approach stems from the fact that for the later waves we condition on a large set of covariates as all intermediate covariates at each node are included. To check whether our propensity score estimates are suffering from overfitting we follow Lechner (2008) and condition on a smaller set of covariates capturing the most recent information and limited information (socio-economic status and health outcomes) from the previous decision nodes. The results from this alternative specification which are presented in Online Supplement Table O6.A2.2 are similar to our main analysis. Additionally, we check whether our results are sensitive to more stringent regression adjustment by conditioning on the covariates from all previous waves. This does not substantially alter our estimates for the mental health effects (Online Supplement Table O6.A2.3). In the alternative models, however, we observe that the high-intensity PCS estimates are most sensitive to model specifications leading to slightly different effect sizes but qualitatively similar results.

7 Discussion & Conclusion

Providing informal care can have negative health effects for informal caregivers. Based on the current literature there is an insufficient understanding of how these effects persist over time and differ by care-intensity and duration. We try to explore these questions by estimating the long-term and dynamic effects of caregiving on caregivers' health using a general population panel survey from the UK.

While early studies on cross-sectional data commonly report caregivers to have low physical health (Carretero et al., 2009), we only find mixed evidence for a causal relationship. Our estimates indicate that informal care leads to a small and short-lived increase in physical health among caregivers providing less than 20 hours of weekly care. A potential alternative explanation for this finding could however be that self-reported physical health is prone to bias as caregivers might change their opinion about their own health by taking the health of the care-recipient as a reference point (Di Novi et al., 2015). For caregivers' mental health outcomes, we find immediate and persisting negative effects of providing care. These effects are heterogeneous and mostly incurred by individuals providing more than 20 hours of care per week. The initial negative effects on mental health slowly decrease in size throughout the years but remain persistent up to four and five years after initial care provision depending on the specification. These effects are, potentially due to limited attrition, more persistent than estimates from previous studies that only found direct effects (De Zwart et al., 2017) or effects up to the first three years of care provision (Schmitz & Westphal, 2015). Additionally, our estimates may be downward biased as high-intensity caregivers are underrepresented in our sample.

For these high intensity caregivers (individuals providing ≥20 hours a week) the estimated negative health effects are similar in magnitude compared to earlier results by Schmitz & Westphal (2015) who focus on individuals providing at least three hours of care on a weekday (≥15 hours). For intensive caregivers the results hence seem robust across different countries with different care systems. This finding also supported by a recent study from Bom and Stöckel (2021) who explore the health effects of informal caregiving for a sample of older UK and Dutch caregivers. For low intensity caregivers this is however not the case: Schmitz & Westphal already find a strong negative effect of -1.9 on the MCS for individuals providing one hour of care per weekday, whereas we do not observe health effects for individuals providing less than 20 hours of care per week in the UK. There might be several explanations for this difference. The intensity levels of care provision are, first, not completely similar and the composition of caregivers within these groups might differ. Second, country differences in the long-term care system and support options might drive changes in the size of the caregiving effect as well as they influence both the selection into care as well as the caregiving experiences within these groups. Courtin et al. (2014) for example describe financial support policies for high-intensity informal caregivers to be available to a wider group in Germany than the UK, mainly due to the strict eligibility criteria in the latter. At the same time there exists a broader range of non-financial support open to all caregivers in the UK compared to Germany. However, given the consistent absence of reliable information on social care consumption in many panel surveys we cannot explore this directly.

Lastly, our dynamic matching results provide insights into the extent to which the static results, indicating decreasing mental health effects over time, are representative for the population of individuals that provide care for multiple consecutive years. There is evidence that the static results do not sufficiently capture that among individuals who provide care for more than one year the mental health effects do not improve over time. Rather for these multi-year caregivers' mental health remains to be negatively affected.

Our study also has several limitations. First, one might question the use of self-reported health measures and prefer, in our case unavailable, administrative information like medical claims or admission data. We believe that given the population we are studying, informal caregivers, these self-reported health measurements better capture changes in health than information regarding health care usage. For mental health this is especially the case as often not all individuals suffering from mental health problems receive or seek treatment. Additionally, administrative information can only capture actual consumption but highly burdened caregivers might forego medical care. Foregoing care could be directly caused by the intensity of caregiving as well as the potential stigma associated with seeking help as a caregiver itself. In addition, our results remain unchanged when using alternative outcome measures (see Online Supplement Figure O6.A2.5) and indicate that the reported effects are economically relevant from the individuals' perspective. This leaves us confident that the reported mental health effects are of interest to policymakers wishing to assess the extent of spillover effects arising from the reliance on informal care to meet social care demands.

While our dataset allowed us to explore the health effects of informal care provision along multiple dimensions not all desired information is available which is a limitation of this study. Firstly, our measure of care intensity is self-reported caregiving hours. Increased hours are likely to reflect a larger overall caregiving burden, however, the tasks performed by caregivers are highly disease-specific and play an important role in the experienced caregiving burden (Pearlin et al., 1990). In addition, Urwin et al. (2021), who use data from USoc from 2015-2017, provide evidence that hours of care provided are reported differently by caregivers and recipients raising questions with respect to what is counted in these self-reported hours. Therefore, reported hours are an incomplete measure. A related cause for uncertainty is the absence of information on why informal care was taken up and discontinued, a process that itself could affect especially mental health outcomes. Another concern refers to our focus on informal caregiving irrespective of whether this occurs alongside formal care as the USoc does not capture such services consistently. Therefore, we cannot explore to what extent these services might serve as a complement or substitute to informal care or help to mitigate the negative health effects in the medium and long run.[18] Ideally future research would have insight into the type of caregiving tasks, formal care use and information on reasons for care take up and discontinuation.

---

[18] USoc wave 7 did include a detailed survey module on informal and formal care sources for recipients; 47% report informal care as the only source of care with 44% reporting a mix of formal and informal care and 8% formal care only. Hours of informal care received are highly similar irrespective of whether it is provided alongside formal care or not. Detailed results are available upon request.

Moreover, despite using a rather large dataset, sample size issues still remained a problem for identification of various additional models we would have liked to run. First of all, the dynamic models showed to be highly data-demanding and we observed too little variation among high-intensity caregivers at later caregiving nodes to be able to identify health effects of providing care for a third year. Second, our dataset lacked information on the health outcomes of many of the care recipients, thereby making our sample too small to separately estimate the family effect, the impact of a family member becoming ill on one's own health. Capturing health outcomes of family members within the survey or the possibility to link data to administrative records to observe the health of family members (as done in Bom et al. (2019b)) would aid in disentangling both effects.

Lastly, an important limitation of our study is its reliance on a matching-based identification strategy and the underlying assumptions. While we do test the robustness of our results to violations of the main identifying assumption and the choice of matching approaches an ideal strategy would rely on exogenous variation in informal caregiving that addresses concerns about time-varying unobserved confounders. Earlier studies exploring the short-term health effects of informal caregiving often relied on parental health shocks or the number of siblings as an instrumental variable. However, the exogeneity of parental health shocks is at least questionable (Schmitz & Westphal, 2015). Recently, Eibich (2021) evaluated the validity of many commonly used instruments confirming these suspicions. Another general concern with respect to alternative IV-based methods is whether the estimated local average treatment effects can be generalized for the entire treatment population (Angrist & Imbens, 1995). Recent studies such as Bakx et al. (2020) or Fischer & Müller (2020) exploited institutional rules and reforms in countries' LTC-sectors as a source of credible exogenous variation in the uptake of formal and informal care use. In our case, however, such an identification strategy was not feasible.

To conclude, our results confirm previous studies reporting negative mental health effects of informal care provision and show that the effects persist up to four or five years after initial care provision. Our estimates suggest that most UK caregivers do not experience substantial adverse health outcomes after providing informal care. However, especially high-intensity caregivers show to be most strongly affected by informal caregiving. This group of caregivers is also most likely to provide care for multiple years which our results suggest to lead to continuing lower mental health. Given the increasing reliance on informal care, these results provide useful insights for policymakers facing difficult trade-offs regarding the allocation of limited resources to support caregivers. Our results indicate that especially high-intensity and long-term caregivers should be targeted to offset the substantial negative mental health effects. While informal care provides undisputable benefits to public health care systems and care-recipients the consequences for those providing the care need to be accounted for.

References

Age UK (2019a). Paying for homecare. https://www.ageuk.org.uk/information-advice/care/paying-for-care/paying-for-homecare/

Age UK (2019b). Paying for a care home. https://www.ageuk.org.uk/information-advice/care/paying-for-care/paying-for-a-care-home/

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430), 431-442.

Barnay, T., & Juin, S. (2016). Does home care for dependent elderly people improve their mental health? *Journal of Health Economics*, 45, 149-160.

Bauer, J. M., & Sousa-Poza, A. (2015). Impacts of informal caregiving on caregiver employment, health, and family. *Journal of Population Ageing*, 8(3), 113-145.

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *The Stata Journal*, 9(4), 524-546.

Bobinac, A., Van Exel, N. J. A., Rutten, F. F., & Brouwer, W. B. (2010). Caring for and caring about: disentangling the caregiver effect and the family effect. *Journal of Health Economics*, 29(4), 549-556.

Bakx, P., Wouterse, B., van Doorslaer, E., & Wong, A. (2020). Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival. *Journal of Health Economics,* 102354.

Bom, J., Bakx, P., Schut, F., & van Doorslaer, E. (2019a). The Impact of Informal Caregiving for Older Adults on the Health of Various Types of Caregivers. *The Gerontologist*, 59(5), e629-e642.

Bom, J., Bakx, P., Schut, F., & van Doorslaer, E. (2019b). Health effects of caring for and about parents and spouses. *The Journal of the Economics of Ageing*, 14.

Bom, J., & Stöckel, J. (2021). Is the grass greener on the other side? The health impact of providing informal care in the UK and the Netherlands. *Social Science & Medicine*, 269, 113562.

Carretero, S., Garcés, J., Ródenas, F., & Sanjosé, V. (2009). The informal caregiver's burden of dependent people: Theory and empirical review. *Archives of gerontology and geriatrics*, 49(1), 74-79.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72.

Carrino, L., Nafilyan, V., & Pabon, A. O. (2019). Should I Care or Should I Work? The Impact of Working in Older Age on Caregiving. *Health, Econometrics and Data Group (HEDG) Working Papers*, 19, 23.

Carers UK (2016). State of Caring 2016. https://www.carersuk.org/for-professionals/policy/policy-library/state-of-caring-2016

Coe, N. B., Guo, J., Konetzka, R. T., & Van Houtven, C. H. (2019). What is the marginal benefit of payment-induced family care? Impact on Medicaid spending and health of care recipients. *Health Economics*, 28(5), 678-692.

Coe, N. B., & Van Houtven, C. H. (2009). Caring for mom and neglecting yourself? The health effects of caring for an elderly parent. *Health Economics*, 18(9), 991-1010.

Colombo, F., Llena-Nozal, A., Mercier, J., & Tjadens, F. (2011). Help Wanted? Providing and paying for long-term care. OECD Health Policy Studies, OECD Publishing.

Comas-Herrera, C., Pickard, L., Wittenberg, R., Malley, J., & King, D. (2010). The long-term care system for the elderly in England. European Network of Economic Policy Research Institutes. ENEPRI RESEARCH REPORT NO. 74 CONTRIBUTION TO WP1 OF THE ANCIEN PROJECT

Courtin, E., Jemiai, N., & Mossialos, E. (2014). Mapping support policies for informal carers across the European Union. *Health policy*, 118(1), 84-94.

Curry, N., Schlepper L and Hemmings N (2019). What can England learn from the long-term care system in Germany? Research report, Nuffield Trust.

Di Novi, C., Jacobs, R., & Migheli, M. (2015). The quality of life of female informal caregivers: From Scandinavia to the Mediterranean Sea. *European Journal of Population*, 31, 309–333.

Eibich, P. (2021). Care or self-care? The impact of informal care provision on health behaviour (No. WP-005-2021). Max Planck Institute for Demographic Research, Rostock, Germany.

European Commission (2018). Informal care in Europe: Exploring Formalisation, Availability and Quality.

ESS Round 7: European Social Survey Round 7 Data (2014). Data file edition 2.2. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

Fernández, J.-L., Forder, J., Truckeschitz, B., Rokosova, M., & McDaid, D. (2009). How can European states design efficient, equitable and sustainable funding systems for long-term care for older people?, Policy Brief No. 11, World Health Organisation Europe, Copenhagen.

Fischer, B., & Müller, K. U. (2020). Time to care? The effects of retirement on informal care provision. *Journal of Health Economics*, 102350.

Glendinning, C. (2013). Long Term Care Reform in England: A Long and Unfinished Story in Reforms. In Ranci, C. & Pavolini, E. (Ed.) *Long-Term Care Policies in Europe.* New York: Springer.

Haley, W. E., & Pardo, K. M. (1989). Relationship of severity of dementia to caregiving stressors. *Psychology and Aging*, *4*, 389-392.

Hirst, M. (2005). Carer distress: a prospective, population-based study. *Social Science & Medicine*, 61(3), 697-708.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 1-24.

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity?. *Journal of Applied Econometrics*, 23(3), 305-327.

Lacey R.E., McMunn A., & Webb, E. (2019). Informal caregiving patterns and trajectories of psychological distress in the UK Household Longitudinal Study. *Psychological Medicine*, 49, 1652–1660.

Lechner, M. (2009). Sequential causal models for the evaluation of labor market programs. *Journal of Business & Economic Statistics*, *27*(1), 71-83.

Lechner, M. & Miquel, M. (2010). Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions. *Empirical Economics*, 39(1), 111-137.

Legg. L., Weir, C.J., Langhorne, P., Smith, L.N., Stott, D.J. (2013). Is informal caregiving independently associated with poor health? A population-based study. *Journal of Epidemiol Community Health*, 67: 95–7.

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Boston College Department of Economics, Statistical Software Components.

Lilly, M.B., Laporte, A. & Coyte, P.C. (2007). Labor market work and home care's unpaid caregivers: A systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work. *Milbank Quarterly*, 85, pp. 641-690.

de la Maisonneuve, C., & Martins, J. O. (2015). The future of health and long-term care spending. *OECD Journal: Economic Studies*, 2014/1, 61-96.

NHS (2018).When the council might pay for your care. https://www.nhs.uk/conditions/social-care-and-support-guide/money-work-and-benefits/when-the-council-might-pay-for-your-care/

OECD (2019). *Health at a Glance 2019: OECD Indicators*, OECD Publishing, Paris.

ONS (2013). 2011 Census analysis: Unpaid care in England and Wales, 2011 and comparison with 2001. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/articles/2011censusanalysisunpaidcareinenglandandwales2011andcomparisonwith2001/2013-02-15#national-comparisons

ONS (2015). UK Health Accounts reference tables. - Long-term care expenditure by function and financing scheme in 2015. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2015

ONS (2017). Unpaid carers provide social care worth £57 billion. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/articles/unpaidcarersprovidesocialcareworth57billion/2017-07-10

Pearlin, L. I., Mullan, J. T., Semple, S. J., & Skaff, M. M. (1990). Caregiving and the stress process: An overview of concepts and their measures. *The Gerontologist*, 30(5), 583-594.

Pinquart, M., & Sörensen, S. (2003a). Associations of Stressors and Uplifts of Caregiving With Caregiver Burden and Depressive Mood: A Meta-Analysis, *The Journals of Gerontology: Series B*, 58(2), 112–128.

Pinquart, M., & Sörensen, S. (2003b). Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. *Psychology and Aging*, 18(2), 250-267.

Rellstab, S., Bakx, P., Garcia-Gomez, P., & Van Doorslaer, E. (2020). The kids are alright-labour market effects of unexpected parental hospitalisations in the Netherlands. *Journal of Health Economics*, 69, 102275.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, 11, 550–560.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.

Rubin, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29, 185-203.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688-701.

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74, 318-328.

Schmitz, H., & Westphal, M. (2015). Short-and medium-term effects of informal care provision on female caregivers' health. *Journal of Health Economics*, 42, 174-185.

Schmitz, H., & Westphal, M. (2017). Informal care and long-term labor market outcomes. *Journal of Health Economics*, 56, 1-18.

Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1.

Townsend, A., Noelker, L., Deimling, G., & Bass, D. (1989). Longitudinal impact of interhouse caregiving on adult children's mental health. *Psychology and Aging*, 4, 393-401.

UK Government (2020). Carer's allowance. https://www.gov.uk/carers-allowance

University of Essex. Institute for Social and Economic Research. (2019). *Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. *12th Edition.* UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-13

Urwin, S., Lau, Y. S., Grande, G., & Sutton, M. (2021). The extent and predictors of discrepancy between provider and recipient reports of informal caregiving. *Social Science & Medicine*, 113890.

Van Houtven, C. H., & Norton, E. C. (2004). Informal care and health care use of older adults. *Journal of Health Economics*, *23*(6), 1159-1180.

Van Houtven, C. H., & Norton, E. C. (2008). Informal care and Medicare expenditures: testing for heterogeneous treatment effects. *Journal of Health Economics*, 27(1), 134-156.

Vlachantoni, A., Robards, J., Falkingham, J., Evandrou, M. (2016). Trajectories of informal care and health. *SSM—Population Health*, 2: 495–501.

Ware, J.E., Kosinski, M., & Keller, S.D. (1995). SF-12: How to score the SF-12 physical and mental health summary scales (2nd. ed). Boston, Massachusetts, The Health Institute, New England Medical Center

Yeandle, S. (2016). Caring for our carers: An international perspective on policy developments in the UK. *Juncture*, 23(1), 57-62.

de Zwart, P. L., Bakx, P., & van Doorslaer, E. K. A. (2017). Will you still need me, will you still feed me when I'm 64? The health impact of caregiving to one's spouse. *Health Economics*, 26, 127-138.

7

CHAPTER 7

General Discussion

Health as an economic good is characterized by its inherent nature as a finite resource with crucial importance for individual-level flourishing. Being in good health, all else equal, enables individuals to participate in economic activities and to enjoy life to the highest attainable level and is a direct contributor to well-being. At the same time, to a certain degree, declining health over the life course is inevitable, with a definitive eventual endpoint but an uncertain course towards it. The decline can be sudden or gradual, permanent or temporary, and is sometimes entirely preventable or at least delayable. The consequences of individual-level health changes when they occur are inherently multi-dimensional. At the individual level declining health has implications for well-being, while the experience of a sudden health shock or a new diagnosis can affect decision-making on investments in health prevention. While these direct effects of health changes are primarily felt by those whose health declines, they also create potential spillovers within their social network. Informal care within and between households might mitigate the consequences of health changes while at the same time the observation of these changes itself could alter the calculus about how we think and invest in our own health even when not directly experiencing a health change (yet). Lastly, at the societal level, the rising share of older individuals in the population experiencing age-related health changes translate into an increasing demand for healthcare, requiring societal trade-offs in its resource-demanding supply.

The chapters of this dissertation studied this multi-dimensional impact of declining health using survey-based subjective outcome measures to illustrate how these measures can help inform economic decision making on health. It did so by focusing on three topics. First, it considered how general population data on subjective well-being can inform health economic evaluations when used to estimate the monetary equivalent value of health changes. The second topic considered the role of health changes in shaping subjective well-being and health perceptions of individuals, and whether these changes are accompanied by corresponding behavioural adjustments. Lastly, the third topic considered the spillovers health changes can have in the wider social network of patients. It did so by exploring the role of health shocks as potential drivers of behavioural change within social networks and the impact of intra- and inter-household informal care provision on caregivers' own health outcomes. This last chapter summarises the main findings on these three topics before providing a discussion of related strengths and weaknesses and broader implications drawn from the findings presented in this thesis.

Synthesis and Discussion of Research Findings

Topic 1: Valuing Health Changes using Survey Data

Most countries with ageing populations are faced with the demographic reality that the share of older individuals in their population increases. As a result of this the coming decades are predicted to see an unprecedented increase in the demand for healthcare. The resources available to satisfy this demand are, however, not endless and decision-makers face competing demands across policy domains from agriculture to the zero-emission transition. These trends underline the need to consider which healthcare needs can and cannot be met from public funds and whether the associated health gains are worth their (opportunity) costs. Economic cost-benefit analyses that translate health gains and losses into monetary equivalents can help identify relevant trade-offs, thereby allowing for welfare improving decisions that consider implicit and explicit alternative spending opportunities. However, making such decisions also requires an answer to a fundamental question: What is the value society attaches to improving population health and how can we express it in monetary terms?

Estimating the monetary value of health remains difficult as a range of competing methodological approaches exist. Most of these approaches rely on the direct elicitation of preferences through choice experiments or contingent valuation exercises. Chapter 2 narrowed down on an alternative approach, the well-being valuation method. Using survey data from Germany it exploited longitudinal variation in health, income, and subjective well-being to estimate the monetary equivalent compensating the loss of one year living in full health. For the average individual in Germany this compensating monetary value was estimated to lie between €20,000 to €60,000 (in inflation-adjusted 2018 prices). The Chapter illustrated that wellbeing-valuation methods using data on subjective well-being, measured using life satisfaction, can provide a complementary estimate to existing approaches used to estimate the monetary value of health. The obtained estimates were in concordance with the previous application of this method to Australian data and, despite being based on a conceptually different approach, they also aligned with estimates produced using alternative methodologies like contingent valuation and healthcare-opportunity-cost-based thresholds.

The well-being valuation approach, however, does have limitations. The range of €20,000 to €60,000 for the estimated population-wide average already highlights the wide variance of monetary values that are obtained from the same dataset but when applying different methodological choices. To obtain a monetary equivalent compensating the loss of one year in full health requires an estimate of the causal effect of income and health changes on subjective well-being. Estimating such causal effects, however, is difficult. Without addressing endogeneity concerns or the possibility of the well-being effects of income to vary along the distribution of earnings, the estimated monetary equivalents were unreasonably large due to the well-documented under-estimation of the effect of income changes on life satisfaction. With respect to the impact of health on well-being, similar concerns remain while the endogeneity of health changes is even harder to address empirically (see the discussions in Chapters 3 and 4). Other practical and conceptual concerns discussed included how the different dimensions of health (e.g., physical, or mental functioning) are aggregated and whether transitioning into health states itself changes the relationship between income and well-being. All of these have implications for the applicability of the obtained monetary equivalents. And, as in the case of experimental elicitation methods, the methodological choices were found to have played a large role in shaping the range of obtained estimates.

Beyond the importance of the underlying assumptions necessary for the quantitative analysis of survey data, Chapter 2 also highlighted the importance of heterogeneities in the underlying population samples. The estimated monetary value of health varied across gender and age-groups but also across the regions in Germany and even based on which years of data collection were considered. These heterogeneities originated from the differential weight that subgroups of individuals attach to income and health as contributors to life satisfaction, which are weighed against each other in the wellbeing-valuation approach, and how broader economic conditions might shape these relationships. Motivated by these heterogeneities and the importance of methodological choices, a range of practical recommendations for applied research were discussed. The results presented in Chapter 2, therefore provide a cautioning example against the broad application of obtained monetary estimates without considering the application context with respect to the patient population, the applicability of underlying assumptions, and the normative implications of using average or subgroup specific values.

Topic 2: The Impact of Health Changes on Perceptions and Behaviours

Adaptation in Subjective Well-being and Health

As outlined in Chapter 2, survey data can be used to quantify the impact of changing health on well-being. Subjective well-being indicators, like life satisfaction, promise to capture the experienced utility of individuals. Compared to indicators of health that focused on specific (physical) health domains, they can provide a more inclusive perspective on the experienced burden arising from declining health. At the same time, precisely due to their subjective nature, such measures are susceptible to phenomena undermining their validity across contexts. One such phenomenon is adaptation, which, broadly defined, describes the process that individuals over time revert to levels of subjective quality of life closer to those experienced before a life event occurred, even if the underlying health issues remain or become more severe. For individuals, adaptation could imply the rational adjustment of expectations and aspirations about what a satisfying life entails to the constraints of declining health. For researchers aiming to quantify the impact of changing health on subjective well-being this poses challenges due to its implications for how gains from medical interventions should be aggregated and compared across populations when adaptation occurs.

To explore the existence and magnitude of adaptation, Chapter 3 used longitudinal survey data from the UK to document how self-reported quality-of-life measures change around the onset of a long-term illness and disability. It documented that life satisfaction deteriorates with the onset of disability but also that reporting patterns trend towards their pre-onset levels already within a few years. This occurs despite individuals' health remaining low or even decreasing further over time. Even when more severe health changes were considered, like those resulting in functional limitations inhibiting day-to-day activities, the pattern of adaptation remained persistent. While these patterns were highly similar across subgroups there was heterogeneity with respect to the magnitude of the observed reverting tendencies. Men and women experienced similar levels of well-being loss over time but the same does not hold if health changes were experienced at different points across the life-course. Individuals aged 55 years and older experienced comparatively small losses in life satisfaction with the onset of disability, while for younger individuals these losses were larger and more persistent. These heterogeneities underline the need to remain cautious when quantifying the impact of ill health over time. They also highlight a trade-off between the additional perspective gained from well-being outcomes and the resulting complexity with respect to their comparability across health states and populations. Whether or not adaptation should be accounted for when comparing the needs of different patient populations, therefore, remains a thorny question, raising both practical as well as normative concerns in the context of health economic evaluations (e.g., Brazier at al. 2018).

Broader well-being measures are not the only subjective outcome measures that could be affected by adaptation. Chapter 3 further documents a persistent pattern of adaptation in self-assessed health, a measure of how individuals perceive their own health to be. Unlike for the broader well-being measure of life satisfaction, however, this does not occur to a degree that indicates complete adaptation. Further, barely any heterogeneity in these patterns is observed across subgroups by age and gender. These differential patterns across measures of subjective health and well-being provide insights into the varying importance (physical) health and the perception thereof has across age-groups. Older individuals are adjusting their health perceptions to a deteriorating health state to a similar degree as younger individuals, but broader well-being measures do not indicate that well-being losses are comparable between these two groups. This suggests that what constitutes broader well-being, and how much weight individuals' perceptions of their own health contribute to this, varies across the life-course.

The Impact of Health Shocks on Health Perceptions and Behaviours

Chapters 2 and 3 of this dissertation focused on questions surrounding the usage of survey-based subjective outcome measures as a quality-of-life measure and potential maximand in the context of health economic evaluations and health policy. Beyond being a convenient tool for researchers exploring the subjective impact of changing health, such measures also capture how individuals themselves perceive their individual circumstances. While economists have often questioned the usefulness of such measures, there is persistent evidence for their predictive power across domains (see e.g., Kaiser & Oswald, 2022), suggesting that they should also provide insights into individual-level decision making. Self-assessed health is a classic example as it represents the perception individuals have about their own health based on their current knowledge about existing conditions and behavioural risk factors, like a smoking habit. However, perceptions are not always accurate. Overconfidence, a lack of health-related subject knowledge, or the inability to interpret new information about one's own health can all result in inaccurate health perceptions. Biased health perceptions here means that conditional on objective health outcomes (e.g., functional limitations that individuals have or pre-existing chronic conditions) and other relevant personal characteristics like age, perceived health is differing from what would be expected given these characteristics.

In this context, health shocks present an interesting case to consider due to their role as an information shock with the potential to result in long-term adjustments to perceived health and behaviours. If health events can at least partially be attributed to behavioural risk factors, like smoking or drinking, they might present a salient reminder on the potential consequences of such behaviours. At the same time, the health trajectory after the shock itself might be amenable to sustained individual-level behaviour change and adherence to medical advice. From this perspective, adaptation in self-perceived health (as documented in Chapter 3) could be seen as a form of biased health perception. Over time, perceptions revert to a baseline level and no longer seem to reflect the health shock and the objective health changes experienced, or the acquired information about their future health risks.

Exploring how perceptions on health are formed following changes to objective health is, however, challenging. Chapter 3 considered the onset of disability as a convenient yet limited example of a consequential health change. Without detailed information on whether disability-onset is caused by gradual decline or a specific disease this leaves open a range of interpretations that could explain the observed pattern in how individual-level perceptions change over time. To explore how health beliefs are influenced by changes in objective health Chapter 4 instead turns towards a natural experiment: the onset of sudden and severe health shocks in the form of a heart attack or a stroke. Both health events have implications for long-term health trajectories and provide information on pre-existing and future risk-factors amenable to behaviour change. Experiencing such shocks, therefore, should not only impact objective health but also have implications for individual-level health perceptions and health-related behaviours due to potential learning effects.

To explore the causal effect of health shocks on health perceptions, objective health, and health behaviours, Chapter 4 combined survey data from the Netherlands with rich administrative records on health and healthcare use at the individual level. The occurrence of a heart attack or stroke leads to a strong decrease in objective health. This is observed both with respect to the risk of future health shocks, in the form of hospitalisations, and the latent health of individuals themselves in their day-to-day activities measured using functional limitations. With respect to perceived health, however, the observed effects over time are different. Instead of trailing the changes in latent health or attenuating over time, indicating a pattern consistent with the adaptation hypothesis (Chapter 3), the effects remain largely unchanged, indicating strong and persistent downward adjustments in perceived health. Further, these results remain virtually unchanged when accounting for latent health directly and the health information patients possessed at the time of the heart attack or stroke. These results imply that following severe health events patients persistently altered their health perceptions conditional on objective health changes, indicating substantial learning effects that go beyond the changes to latent health.

Chapter 4 further documented that in line with persistent learning effects, substantial behavioural adjustments in health-related decision making occur among patients as they increase health prevention efforts. The prevalence of risky health behaviours like smoking and excessive drinking decreases substantially immediately after a health shock occurs, with the smoking rate among heart attack patients, for example, dropping by 14 percentage points or 46% of the 30% smoking rate observed among heart attack patients the year before their heart attack. More importantly, however, these improvements in health prevention efforts remain persistent for seven years after the event with no evidence of attenuation. Similarly, persistent effects were observed in the uptake of preventive medication use to decrease the risk of future cardiovascular events, which is near universal with no sign of adherence decreasing over time. The combined findings on health perceptions and behaviours documented in this chapter underline the importance of the time immediately after a health shock as a window of opportunity for behavioural interventions aimed at improving health behaviours among vulnerable populations. They also suggest that although persistent health behaviour changes are notoriously hard to achieve, the key to their success might be found in how individuals perceive their own health to be and in identifying ways to persistently influence these perceptions.

Topic 3: Spillover Effects of Declining Health

Health Shocks and Spouses' Perceptions and Behaviours

Declining health at the individual level is societally relevant due to its implications for healthcare use in the context of ageing populations and resource scarcity. However, already at the level of social networks, individuals' health shocks can have second order effects. Chapter 4 documented how health shocks shape health perceptions and behaviours of patients but the potential informational value of such events is not limited to patients alone. Spouses or other family members sharing similar health-related habits, pre-existing conditions or potential genetic risk factors might update their own health perceptions as they learn about shared risk factors, or simply due to being confronted with a salient signal prompting them to think about health in general. Such behavioural spillovers have been explored in a range of contexts (e.g., Fadlon & Nielsen, 2019), but it remains unclear whether behavioural response is a direct reaction to a highly salient event in their social environment or also because family members themselves adjust their own health perceptions.

The second half of Chapter 4 explored how heart attacks and strokes affect the perceptions and behaviours of spouses of patients. As one would expect we do not find the same level of effects as for the patients who experienced the health event. However, contrary to existing findings in the literature, we find only limited evidence for any spillovers from heart attacks and strokes on the perceptions and behaviours of spouses. While there is some limited evidence for these events also impacting the perceived health of spouses and lowering the likelihood to smoke, the observed effects show a considerable degree of noise. Further, we do observe strong effects on preventive medication use, one area in which previous studies found some of the strongest spillovers reaching as far as close co-workers (Fadlon & Nielsen, 2019). However, these effects can largely be attributed to pre-existing trends in prescription medication use across different birth-cohorts among our sample of spouses. Therefore, we find little evidence for health events to result in significant behavioural spillovers or learning effects among spouses. It needs to be kept in mind, however, that these results could be driven by specific methodological choices, in particular the focus on a small subsample of the general population of patients of spouses, and therefore should be interpreted with caution.

Health Shocks and Informal Care in the Household and Family Network

Health shocks to family and household members can have informational value but also result in additional strains on shared resources of the household or the broader family network. The most direct source of strain is the (temporary) need for additional support to allow them to participate in society and maintain their quality of life after experiencing a health change. Providing this support is generally referred to as social care or non-health related care, which is often only partially delivered by public providers. In many high-income countries, even those with comparatively generous public healthcare systems, family members (primarily daughters and female spouses) provide most of these services in the form of informal care. As long-term care services have seen an increasing trend towards ageing-at-home policies or related initiatives aimed at cutting the usage of costly nursing home care, this demand for unpaid informal care will likely increase. Providing such informal care can be physically and mentally demanding, raising the question to what extent and increasing reliance on inter- and intra-family care to meet social care demand generates health losses for caregivers themselves.

Chapter 5 used data from the Netherlands and the United Kingdom, two countries with substantially different long-term-care systems, to explore whether providing informal care decreases the health of caregivers themselves. Despite considerable differences in the generosity of public care provision in both countries, a similar picture emerged. While informal care provision had little direct effect on physical health, high-intensity caregiving (i.e., more than 20 hours a week) had negative effects on mental health outcomes of caregivers. These results highlight the importance of the overall healthcare system characteristics in determining the population-level impact of informal caregiving through its role in shaping the caregiving population. Conditional on the intensity of care provided, the mental health impacts of providing informal care might be broadly comparable even across dissimilar long-term care systems. However, health system characteristics do shape how many individuals provide care at high intensity levels, with less generous systems as found in the UK resulting in a larger share of high intensity caregivers. The analyses presented in Chapter 6 further illustrate that these effects are persistent over multiple years and larger for those caregivers that provide care for longer periods of time. This type of care is especially provided by female spousal caregivers and likely tied to the type of health changes experienced by the care recipients. The long-term declines in the health of older individuals can therefore cause spillovers on the mental health of caregiving spouses and children. These spillovers should be accounted for when considering policy interventions that can decrease the care-dependency of older individuals, as the sole focus on direct benefits to recipients or fiscal resources would underestimate their overall benefits.

Strengths and Limitations

This dissertation combines a broad range of methodological approaches to study the impact of changing health on individuals and their families. It does so to quantify the impact of deteriorating health across multiple dimensions and to illustrate the advantages and disadvantages of data on survey-based subjective outcome measures in this context. The individual chapters discuss in detail the merits and limitations of the respective methodologies applied. However, as all chapters share a common methodological approach it is worthwhile to discuss overarching strengths and limitations resulting from this.

This dissertation, in line with broader trends in economics and other quantitative social sciences, is based on the analysis of large-scale, high-quality survey data in combination with modern econometric methods aimed at causal inference. As such it follows the best practices of what has been called the "credibility revolution" (Angrist & Pischke, 2010), which has influenced empirical research in economics and other quantitative social sciences towards policy-oriented questions answered using robust, replicable, and transparent research methods. This aspiration is reflected across multiple dimensions. Chapter 2 focuses entirely on the discussion of methodological choices in health economics research on the monetary value of health and their consequences for the application of the obtained estimates in real-world contexts. The results on adaptation are based on cross-country evidence from the UK (Chapter 3) and the Netherlands (Chapter 4) while also the results on the health impact of providing informal care are based on these two countries (Chapters 5). Even more so Chapters 3 and 4 also contrast two complementary research designs when exploring how health shocks shape health perceptions. The results of all chapters are based on survey-based subjective outcome measures, but their insights are not simply taken at face-value. Multiple chapters consider the advantages and disadvantages of such measures ranging from methodological choices in their analyses (Chapter 2) to the extent to which subjective outcome measures relate to objective outcome changes over time (Chapters 3 and 4). Throughout all chapters the usefulness of survey-based subjective outcome measures is critically discussed with a re-occurring emphasis on many limitations being potentially overcome by complementary administrative data sources. Chapter 4 directly shows how the availability of linkable survey and administrative register data provides tangible benefits to modern empirical research in this context. Lastly, throughout this dissertation special attention is given to the transparent presentation of the conducted research. For some chapters, the appendix is only an assortment of additional tables and graphs mentioned somewhat in passing. For the majority, however, the appendices contain additional information and their exclusion from the main body of the text should not be interpreted as a definitive judgement on their importance. This effort towards transparency culminates in the publication of research methods and materials in the form of replication packages for the analyses presented in Chapters 2 and 3[19], covering the entire path from raw data to published results. While this is not aligned with open science principles along all possible dimensions, as none of the chapters were accompanied by a pre-registered analysis plan, it is a deliberate step aimed at increasing the reproducibility and accessibility of the presented research.

---

[19] Both can be found on the *Open Science Framework* where the materials for Chapter 2 (https://osf.io/b8nsz/) and Chapter 3 (https://osf.io/q7fx9/) have been published. The replication packages contain all materials necessary to replicate the presented results based on the original unprocessed data used in each chapter, the German Socio-Economic Panel (Chapter 2) and Understanding Society (Chapter 3), which cannot be shared but access to which is free of charge for non-commercial users.

This dissertation aims to follow best practices in modern empirical research in health economics and policy. As it aims to quantify the impact of deteriorating health using existing survey-based subjective outcome measures, the focus on the quantitative analysis of survey data is not surprising but naturally comes at some cost. Throughout this dissertation theoretical perspectives are mostly used as a motivating background rather than a direct focus of the analysis. Except for the discussion of the relation between empirical estimates and their varying theoretical implications and interpretations presented in Chapter 4, they are primarily used as a brief motivation for or against specific hypotheses. In addition, no chapter examines a specific policy change or reform in the context of which estimated effects can be directly connected to a policy-lever such as a change in retirement age or a tightening of eligibility rules guarding access to public long-term care. This hampers the translation of evidence presented in this dissertation into specific recommendations for policymaking. Without a more direct use of theoretical perspectives in combination with empirical methods the underlying mechanisms that drive observed results sometimes remain an object of informed speculation. However, precisely these mechanisms help to ascertain whether the obtained results remain applicable across contexts and their informational value for specific policy designs. Even a precisely identified causal parameter can become meaningless without understanding the underlying mechanisms producing the observed effect or any effort to quantify its economic relevance. Questions of causality take centre-stage in most chapters of this dissertation and modern social science research more broadly, but they are only a necessary and not a sufficient condition for translating empirical results into policy relevant insights.

An additional limitation with respect to the generalisability of findings can be found in the necessary balancing act between fitness for purpose of data sources and credible methodological approaches, two conditions that often cannot be jointly satisfied. Chapter 4 provides an illustrative example. Based on a combination of survey and administrative data from the Netherlands it explores how health shocks (i.e., heart attacks and strokes) affect health perceptions and behaviours. It combines a short general population survey, conducted among more than one million unique respondents, with detailed administrative data on individual-level health changes. This combination allows for the chosen methodological approach to work. Heart attacks and strokes as sudden and unpredictable health events provide a credible source of identifying variation but exploiting this variation to study the impact of health events on survey-based outcomes measures also requires sizeable samples of patients among survey participants. It is this constraint that the data combination overcomes. At the same time, this unique context has implications limiting the broader applicability of findings. Most health shocks experienced by individuals are less severe than heart attacks or strokes, making them an extreme and highly specific case-study. Relying on a measure of overall self-assessed health, on the other hand, represents the broadest possible conceptualisation of health perceptions with alternative measures like subjective survival expectations being more relevant to many questions on health-related economic decision making. Lastly, while the research design allows for a credible identification strategy it is also difficult to emulate other contexts given the limited availability of similar data combinations outside of the Netherlands. This illustrates a common caveat in the analysis of survey-based subjective outcome measures collected using general population samples; they are seldomly designed to answer a single research question or made available in a context that allows researchers to supplement these data with administrative records. More generally, what survey items were included and deemed important at one point in time continues to set the guardrails both with respect to what research questions can be asked and how they are answered. All these aspects are to a certain degree also applicable to the other chapters in this dissertation and underline the trade-offs when balancing fitness for purpose with the practical demands of empirical research.

The results presented in this dissertation also suffer from a clear absence of any results that are not based on quantitative analyses. General population survey data promises to provide insights with a certain level of objectivity, but it has blind spots. Counter-balancing the subjective nature of such outcome measures with the law of large numbers and assumptions underlying econometric analyses always results in constraints. Even when the necessary assumptions are plausible, an isolated silo of information remains, more so as large-scale surveys require a certain level of generalisation to combine different topics from labour market participation to health behaviours. This often results in a combination of short surveys on a variety of topics, allowing many research questions to be addressed but limiting the potential of exploratory depth in specific directions. This further tightens the corridor in which the obtained results can be interpreted. It also highlights the need for complementary insights from experimental or qualitative studies, as they are often pivotal to our understanding of any research question. As such, the research presented in this dissertation and its interpretations in most circumstances cannot stand by itself. The results presented in Chapters 5 and 6 on the health impacts of informal caregiving are exemplary for this. While they both identify the intensity of caregiving, measured using hours of care provided, as an important mechanism driving mental health differences across caregivers, they cannot go beyond this broad classification. As not all caregiving activities are comparable in their time-effort and levels of stress or enjoyment, such a classification provides only a rough guideline for policymakers. This highlights the area in which qualitative small-sample studies allowing for a more open-ended collection of data emphasizing individuals' complex experiences can help add necessary nuance and guide future research. The findings of this dissertation should, therefore, be read in the context of complementary evidence from multi-disciplinary research. Even more so in the light of the related limitations regarding the need to identify relevant mechanisms that underlie the observed empirical results.

A last limitation can be found in the general scope of this dissertation and the degree to which its individual parts are synthesized as part of the substantive chapters. Each of the topics covers a different dimension relevant for quantifying the impact of deteriorating health using survey-based subjective outcome measures. In doing so, it considers methodological aspects of the usage of such data while also highlighting their comparative advantages and disadvantages. It moves from a societal (Chapter 2) to an individual (Chapters 3 and 4) and a household and family-network perspective (Chapters 4, 5, and 6), and explores how observational data on survey-based subjective outcome measures can inform our understanding of health economic decision making within these dimensions. While this provides an overview on the various dimensions that are impacted by deteriorating health, it does not provide a unified framework on how to aggregate any identified effects across dimensions nor quantifies their relative importance within such a framework. Chapters 2 and 3 on the use of broader well-being measures and the discussion of caregiver spillover effects in Chapters 5 and 6 all have links to the question on how to define the maximand when setting health policy priorities with respect to what measure to use and among whom they might be elicited. If deteriorating health at the individual level leads to well-being losses directly and indirectly through the mental health of family members providing care, should these be seen as an integral part of the total cost of deteriorating health? The answer to this question could have tangible implications on how the burden of disease from certain informal care intensive conditions such as dementia is evaluated compared to other diseases not necessarily resulting in such spillover effects (Chandra et al., 2020). How and under which assumptions such decentralized effects could and should be weighted up and aggregated, and where to draw the line on what effects are relevant in the first place, cannot be determined based on the presented results alone. To a certain degree this is justified, as research is not obliged to answer all questions deriving from a particular result, with policy makers bearing the responsibility to synthesize evidence into decisions. Nonetheless, it is a limitation highlighting the continued relevance of empirical and theoretical welfare economics to help formalize decisions on health and healthcare.

Implications for Policy and Future Research

Keeping in mind the strengths and limitations of this dissertation, a set of implications for policy-making and future research directions is evident. The first implications concern the use of survey data for policy-oriented research. This dissertation illustrates how survey-based subjective outcome measures can play a unique role in quantifying the impact of health changes across various contexts. Researchers and policymakers should, however, keep a healthy level of scepticism when interpreting results, including those presented in this dissertation. The analysis of survey data requires a range of choices that can drastically alter the obtained results (Chapter 2), but which are not always transparently discussed, reported, or made available to the public eye to scrutinise. The heterogeneity of results is not only confined to how a specific statistical analyses is conducted (Chapter 2) but also to which type of analyses is chosen (Chapter 3 and 4), and which types of measures, timeframes and populations are considered (Chapter 2 and 3). In this context, the availability of supplementary data sources on more objective outcome measures, for instance, by linking administrative records on healthcare consumption (Chapter 4), provides an invaluable but unfortunately scarce resource for research. For policymakers wishing to make use of the unique perspective that can be obtained from data on subjective outcome measures to the largest possible benefit, increasing the availability of linkable administrative data records should become a key priority.

Irrespective of this cautionary note, there are also many insights to be gained based on survey-based subjective outcome measures. As populations age, physical health becomes a maximand with inherently limited scope for improvement. Subjective well-being and related measures

(Chapters 2 and 3) can provide a more inclusive perspective when determining the burden of ill health empirically. This perspective does not come without its own shortcomings, but it emphasizes aspects that would remain hidden with a narrow focus on objective “hard” outcomes or a unidimensional view on what constitutes good health over the life-course. Survey-based outcomes measures can also help approximate the true scope of affected populations. The case of informal caregiving (Chapters 5 and 6) underscores this. Administrative hospital records and other routinely collected data can capture hip-fractures, falls or other health events prevalent among older populations. These records, however, seldomly capture who might care for an elderly parent after they leave the hospital ward. In most countries, informal care by relatives is observable in administrative records only under restrictive conditions such as the receipt of public long-term care, which limits the representativeness of the caregiving population identifiable based on such data. Only focusing on those whose caregiving context qualifies their inclusion, for example, based on hours of care provided or the care-recipients official care needs and financial assets, would provide a skewed picture. This is a classic example of a data-gap that carefully collected surveys fill. Researchers, and especially economists, should embrace the fact that some questions are important enough to explore even if the data available puts restrictions on the methodological sophistication that can be applied. If research is supposed to inform policy making, it needs to recognize that in many real-world contexts decisions cannot be deferred until the perfect natural experiment has emerged or a measure of revealed as opposed to stated preferences has become available.

Beyond these broader considerations, there is also a set of more concrete implications for future research directions. On the topic of spillover effects for informal caregivers and the use of well-being measures in health economics, policy-making and future research would benefit from a focus on priority setting from a public preference perspective regarding what should and what should not be included in (health) economic evaluations. This holds both with respect to the conceptual and methodological basis of decisions on the allocation of resources in the production of health. While there is a range of estimates available on the partial effects of certain policy alternatives based on different methodological approaches, there is a lack of research on what types of evidence broader members of the public and experts prefer, why they do so, and under what conditions (although recently studies like Dur et al. (2023) have started to explore this in the context of voters’ and politicians preferences on policy experiments). This contrasts with research on conceptual aspects, such as equity related considerations in the allocation of healthcare, which are more commonly explored. As discussed in Chapter 2, obtaining estimates on the monetary value of health remains difficult, with various competing approaches available. However, the degree to which certain conceptual concerns are addressed across methodologies differs, and little systemic evidence exists on how the public, or subject expert groups such as medical doctors or health economists themselves, would rank different methodological alternatives and trade-off their comparative advantages and disadvantages.

With respect to the impact of health shocks on health perceptions (Chapter 4), the implications are more directly informative for research and policy design. Health perceptions seem to be a strong mediator of behavioural adjustments as individuals update their perceptions in a long-term persistent manner after experiencing a severe health shock with high informational value. This reaffirms findings from existing studies from medical and health economic literatures concerning a window of opportunity during which individually tailored behavioural interventions might be particularly promising (Fadlon & Nielsen, 2019). At the same time, it points towards more targeted research necessary to add nuanced understanding of the underlying mechanisms. Even after severe health shocks there are individuals not adjusting their behaviour. The results reported in Chapter 4, for example, imply that only 46% of all smokers suffering from a heart attack quit this risky health behaviour, although almost all take up preventive medications. For some this might simply be a matter of true preferences over the trade-off between health investments and consumption utility from risky health behaviours. For others, it might instead be explained through various mechanisms. Misinterpreting the information received from health shocks or becoming overly fatalistic about post-shock survival could each result in an absence of behavioural adjustments but have different implications for the observed changes in perceived health and health behaviours. Focusing on this direct link between behaviours and different conceptualisations of health perceptions and beliefs provides a fruitful starting point for future research. This is also an area in which more tailored survey measures, even when collected among smaller samples, would provide the crucial insights necessary for efficient policy design. They could help identify among which populations salience or learning mechanisms are dominant drivers of behavioural responses, and what type of public health information interventions could emulate these mechanisms as a tool to improve individual-level health prevention efforts. Given the high cost-effectiveness of certain lifestyle-related health prevention efforts lowering the risk for health shocks (see e.g., Newhouse, 2021), such insights could be highly impactful. A natural extension to this is to further to explore areas of economic decision making done based on health expectations (e.g., insurance choice or retirement decisions). Both are characterised by a high-level of individual responsibility in decision making while at the same time there is an abundance of economically sub-optimal choices being observed among real-world economic agents resulting in tangible welfare losses (e.g., Handel et al., (2020) in the context of healthcare deductible choices in the Netherlands), making it another area in which behavioural insights could be particularly impactful.

An interlocking area of future research with high policy relevance in the context of behavioural spillovers is the healthcare system context in which such spillovers might occur, or not occur. Contrary to what was expected based on existing evidence (by Fadlon & Nielsen (2020), Hodor (2021), and Hoagland (2021)), the results presented in Chapter 4 suggest only limited spillovers from heart attacks and strokes on the perceptions and behaviours of patients. There are potential methodological explanations for this, for example, the chosen sample of patients and their spouses or insufficient statistical power to detect comparatively small effects. However, these results could also imply that spillovers are moderated by broader healthcare system characteristics. Even with the intention to act upon a salient cue the defining constraints are set by the broader healthcare system and how it shapes the information-environment and the choice options presented to consumers. Gatekeepers (e.g., general practitioners and family doctors) might not as readily make available prescription medications but instead offer other healthcare services like screening or routine check-ups instead. This might be a desirable instance of healthcare professionals managing and steering consumption towards high-value care and discouraging wasteful use (Hoagland, 2021). But it might also be a source of health inequalities as it raises the question on how different population groups manage to navigate the healthcare system (Chen et al., 2022; Schwab & Singh, 2023). These areas are particularly interesting to consider in a cross-country framework, to draw conclusions how healthcare policies shape these, but, as outlined above, are also contingent on the availability of high-quality administrative data.

Concluding Remarks

This dissertation began with a bleak outlook on the author's likely health trajectory over the coming decades. How this outlook holds up against realised future developments remains to be seen, but as of the completion of this dissertation his perception of the last years would suggest that he is bucking the trend (temporarily). Some exceptions to the rule, however, do not scale up against the weight of macro-level forces. The coming decades will see an increasing demand for healthcare in high-income economies, driven by the rising number of individuals experiencing declining health themselves or in members of their family. Whether these declines are by virtue of age, bad luck, or (partially) attributable to lifestyle choices does not alter the fact that while the primary consequences are experienced individually, their impact often goes much further. How societies meet future healthcare demand will likely require uncomfortable choices on whose health is prioritised, how to address the increasing importance of preventable conditions, and whether the burden of elderly care is primarily borne by families or the broader public. The chapters of this dissertation explored how survey-based subjective outcome measures can help quantify the impact of declining health across these dimensions to inform choices on these dilemmas. The subjective experiences of patients and their family members or caregivers may not be perfect substitutes for the objective measures health economists often prefer. However, as the chapters of this dissertation have hopefully shown, they do nonetheless play a distinct role in identifying the broader impact of declining health to inform economic decision making.

References

Angrist, J. D., & Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3-30.

Brazier, J., Rowen, D., Karimi, M., Peasgood, T., Tsuchiya, A., & Ratcliffe, J. (2018). Experience-based utility and own health state valuation for a health state classification system: why and how to do it. *The European Journal of Health Economics*, 19(6), 881--891.

Chandra, A., Coile, C., & Mommaerts, C. (2020). What Can Economics Say About Alzheimer's Disease? (No. w27760). National Bureau of Economic Research.

Chen, Y., Persson, P., & Polyakova, M. (2022). The roots of health inequality and the value of intrafamily expertise. *American Economic Journal: Applied Economics*, 14(3), 185-223.

Dur, R., Non, A., Prottung, P., & Ricci, B. (2023). Who's Afraid of Policy Experiments? Retrieved from osf.io/mnjv3

Handel, B. R., Kolstad, J. T., Minten, T., & Spinnewijn, J. (2020). The social determinants of choice quality: evidence from health insurance in the Netherlands (No. w27785). National Bureau of Economic Research.

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences*, 119(42), e2210412119.

Newhouse, J. P. (2021). An ounce of prevention. *Journal of Economic Perspectives*, 35(2), 101-18.

Fadlon, I., & Nielsen, T. H. (2019). Family health behaviors. *American Economic Review*, 109(9), 3162-91.

Hoagland, A. (2021). An ounce of prevention or a pound of cure? The value of health risk information. Working paper.

Hodor, M. (2021). Family health spillovers: evidence from the RAND health insurance experiment. *Journal of Health Economics*, 79, 102505.

Schwab, S. D., & Singh, M. (2023). How Power Shapes Behavior: Evidence from Physicians. Working paper.

A

APPENDICES

Summary

Health is a vital economic resource that is crucial for individual-wellbeing but is characterized by its inherently declining nature over the life-course. Declining health has a range of implications for those who experience it, their families, and society more broadly. Observational survey data on subjective outcomes are an important tool to quantify these broad impacts. Subjective outcome measures concern individual experiences and promise to provide a broader perspective on the impact of declining health than objective health measures like days spent in hospital or disease incidence rates. At the same time, using subjective measures comes with distinct challenges. The primary objective of this dissertation is to investigate how observational survey data can help quantify the broad impacts of declining health. To this end, the chapters of this dissertation cover three topics: *Valuing Health Changes using Survey Data*, *The Impact of Health Changes on Perceptions and Behaviours*, and *Spillover Effects of Declining Health*.

The first topic, *Valuing Health Changes using Survey Data*, explores how broader subjective well-being measures, such as life satisfaction, can inform health economic evaluations as a tool for estimating the monetary value of health changes. To do so, Chapter 2 uses general population survey data from Germany to exploit individual variation over time in life satisfaction, health, and income to estimate the monetary amount that is necessary to compensate for the loss of one year in full health. This study provides the first estimates of the monetary value of a quality-adjusted life-year for Germany and explores several methodological concerns when using this relatively new approach.

The second topic, *The Impact of Health Changes on Perceptions and Behaviours*, narrows down on the stability and comparability of self-reported outcome measures over time, and their relationship with objective health information. Chapter 3 explores whether individuals' self-reported life satisfaction and health adapt to the onset of long-term disability, using longitudinal survey data from the UK. This study documents a consistent pattern of adaptation to disability, especially in measures of broader well-being such as life satisfaction. As subjective measures of health perceptions are also interesting to consider as in input to individual-level decision-making Chapter 4 exploits a unique combination of individual level survey data and administrative health records to analyse the causal impact of health shocks on health perceptions and behaviours over the long run. It shows that experiencing a heart attack or stroke results in persistent adjustments to self-perceived health and increased health prevention efforts by reducing risky health behaviours such as alcohol drinking and smoking.

The third topic, *Spillover Effects of Declining Health*, focuses on spillover effects from individual-level health changes on direct family members. The second half of Chapter 4 explores the impact of heart attacks and strokes on the spouses of patients in terms of their own health perceptions and behaviours. Unlike the large behavioural adjustments observed among patients it documents only limited evidence for behavioural spillovers on the spouses of patients. Chapters 5 and 6 consider the role of inter- and intra-household support as a response to health shocks and the

potential associated spillovers on the health of informal caregivers themselves. Using survey data from the UK and the Netherlands it analyses the short- and long-term health effects of informal caregiving to caregivers and provides causal evidence for substantial and persistent mental health losses among those informal caregivers providing high-intensity care.

The chapters of this dissertation demonstrate the multi-dimensional impact of health changes and how survey-based subjective outcomes measures can help understand this impact and inform health policy. As measures of experienced utility, they can help estimate the monetary value of health changes (Chapter 2) and show how people over time adapt to health changes (Chapter 3). At the same time, when paired with detailed information on objective health changes, measures on self-perceived health can illuminate how individual-level decision making is impacted by health events, and the degree to which persistent learning and behavioural effects occur (Chapter 4). Lastly, survey-based subjective outcome measures can also help identifying the scope of spillover effects in dimensions that are hard to quantify or among populations that are usually not covered by routinely collected administrative data, by providing an estimate of the physical and mental health effects of providing informal care among different populations of family caregivers (Chapters 5 and 6).

Zusammenfassung

Die individuelle Gesundheit stellt eine wesentliche ökonomische Grundlage dar, die nicht nur für das individuelle Wohlbefinden entscheidend ist, sondern auch geprägt ist durch eine natürliche Verschlechterung im Laufe des Lebens. Eine Verschlechterung des Gesundheitszustandes wirkt sich auf die Betroffenen, ihre Familien und die Gesellschaft insgesamt aus. Erhebungen über das subjektive Empfinden spielen eine wichtige Rolle bei der Quantifizierung dieser Effekte. Solche Umfragedaten erfassen die persönliche Erfahrung gesundheitlicher Veränderungen und ihre Konsequenzen im alltäglichen Leben. Dadurch erlauben sie eine umfassendere Sicht auf die Auswirkungen abnehmender Gesundheit als solche Indikatoren, welche sich nur auf die körperliche Gesundheit konzentrieren, wie die Anzahl der Krankenhausaufenthalte oder die Häufigkeit von Erkrankungen. Gleichzeitig stellen die Analyse und Interpretation solcher Indikatoren auch eine besondere Herausforderung dar. Hauptziel dieser Dissertation ist es, die Bedeutung solcher Umfragedaten für die Quantifizierung der multidimensionalen Konsequenzen abnehmender Gesundheit zu ermitteln. Um dies zu erreichen, befassen sich die einzelnen Kapitel mit drei zentralen Themenbereichen: Die Nutzung von Daten über das subjektive Wohlbefinden zur Errechnung monetärer Gegenwerte von gesundheitlichen Veränderungen, dem Einfluss von Gesundheitsschocks auf die Wahrnehmung der eigenen Gesundheit und das Gesundheitsverhalten, und die Erfassung von indirekten Übertragungseffekten.

Der erste Themenbereich untersucht, wie Daten zum subjektiven Wohlbefinden, wie zum Beispiel die generelle Lebenszufriedenheit ökonomischen Abwägungen im Gesundheitssektor informieren können. Zu diesem Zweck werden in Kapitel 2 Längsschnittdaten einer repräsentativen Stichprobe der deutschen Bevölkerung analysiert um den monetären Gegenwert von Gesundheit zu Schätzen. Auf Basis der zeitlichen Veränderungen von Gesundheit, Lebenszufriedenheit und Einkommen und der Zusammenhänge dieser drei Faktoren wird der finanzielle Gegenwert von Gesundheitlichen Veränderungen errechnet. Dadurch liefert diese Studie erste Schätzungen für den monetären Wert eines qualitätskorrigierten Lebensjahres in Deutschland und thematisiert methodologische Herausforderungen, die sich in der Errechnung solcher hypothetischen Gegenwerte stellen.

Der zweite Themenbereich befasst sich mit den Auswirkungen von Gesundheitsveränderungen auf die Wahrnehmung der eigenen Gesundheit und das Gesundheitsverhalten sowie die generelle Lebenszufriedenheit. Kapitel 3 untersucht auf Basis von Daten aus dem Vereinigten Königreich, ob und wie sich Lebenszufriedenheit und das subjektive Gesundheitsempfinden an langfristige gesundheitliche Einschränkungen anpasst. Die Ergebnisse zeigen auf ein persistentes Muster, welches nahelegt, dass durch sich Anpassung über die Zeit sich die Lebenszufriedenheit und das Gesundheitsempfinden wieder an das Level vor einer gesundheitlichen Verschlechterung annähert, insbesondere bei der Lebenszufriedenheit. Da das Gesundheitsempfinden auch eine Rolle in den Entscheidungen von Individuen spielt, zum Beispiel beim Präventionsverhalten, kombiniert Kapitel 4 Umfragedaten mit administrativen Gesundheitsdaten, um den kausalen Zusammenhang von Gesundheitsschocks auf die Wahrnehmung und das Verhalten von

Individuen zu analysieren. Es dokumentiert, dass gesundheitlichen Schocks wie Herzinfarkte oder Schlaganfälle zu dauerhaften Änderungen in der Gesundheitswahrnehmung und zu einer ebenso dauerhaften Zunahme an präventiven Gesundheitsverhalten führen wie der Aufgabe des Rauchens oder einem geringeren Alkoholkonsum.

Der dritte Themenbereich beleuchtet die indirekten Auswirkungen abnehmender Gesundheit auf die Familienmitglieder der Betroffenen. Es zeigt, dass signifikante Verhaltensänderungen hauptsächlich bei den Patienten selbst auftreten, während bei den Partnern von Patienten nur begrenzte Anpassungen festzustellen sind und die Übertragungseffekte in diesem Bereich somit limitiert scheinen. Die Kapitel 5 und 6 erforschen inwieweit Angehörige durch ihre Pflegearbeit selbst Gesundheitsveränderung erfahren, einem indirekten Übertragungseffekt von individueller gesundheitlicher Verschlechterung. Auf Basis von Daten aus den Niederlanden und dem Vereinigten Königreich präsentieren diese Kapitel Evidenz für den kausalen Zusammenhang zwischen dem Leisten von Pflegearbeit und kurz- und langfristiger Verschlechterungen im Bereich der mentalen Gesundheit, vor allem für solche Pflegende welche sich in hoher Intensität um kranke Angehörige kümmern.

Die Kapitel dieser Dissertation illustrieren die vielschichtigen Auswirkungen von Gesundheitsveränderungen und die Bedeutung von Umfragedaten über das subjektive Empfinden für unser Verständnis dieser Effekte und Gestaltung von Gesundheitspolitik. Indikatoren über das subjektive Wohlbefinden können genutzt werden um den monetären Gegenwert von Gesundheit zu Schätzen (Kapitel 2) und zeigen das Individuen sich an gesundheitliche Veränderung über die Zeit anpassen (Kapitel 3). In Kombination mit detaillierten objektiven Gesundheitsdaten können Umfragen über das individuelle Gesundheitsempfinden unser Verständnis über Entscheidungsprozesse zum Präventionsverhalten verbessern und uns Erlauben zu verstehen wie Gesundheitsschocks Lerneffekte und Veränderungen im Gesundheitsverhalten nach sich ziehen (Kapitel 4). Außerdem erlauben uns Umfragedaten die Existenz und Relevanz von Übertragungseffekten zu quantifizieren, vor allem für solche Populationen, welche oft nicht in Administrativen Daten erfasst werden, wie zum Beispiel Pflegende Angehörige, die oft selbst gesundheitliche Verschlechterungen aufgrund ihrer Pflegetätigkeit erfahren (Kapitel 5 und 6).

Nederlandse Samenvatting

Gezondheid is een fundamentele economische bron die essentieel is voor persoonlijk welzijn, maar kenmerkt zich door een natuurlijke afname gedurende de levensloop. Afnemende gezondheid heeft verschillende implicaties voor hen die het ervaren, hun families en de maatschappij als geheel. Observationele gegevens uit vragenlijsten over subjectieve uitkomsten zijn een belangrijk hulpmiddel om deze brede effecten te kwantificeren. Subjectieve uitkomstmaten omvatten individuele ervaringen en bieden een breder perspectief op de gevolgen van afnemende gezondheid dan objectieve maten zoals ziekenhuisdagen of incidentiecijfers van ziektes. Tegelijkertijd brengt het gebruik van subjectieve uitkomstmaten specifieke uitdagingen met zich mee. Het hoofddoel van dit proefschrift is te onderzoeken hoe observationele gegevens uit vragenlijsten kunnen helpen om de brede effecten van afnemende gezondheid te kwantificeren. Om dit te bereiken, behandelen de hoofdstukken van dit proefschrift drie kernthema's: Het waarderen van gezondheidsveranderingen via gegevens uit vragenlijsten, het effect van gezondheidsveranderingen op percepties van gezondheid en gezond gedrag, en de indirecte effecten van afnemende gezondheid.

Het eerste thema, het waarderen van gezondheidsveranderingen via gegevens uit vragenlijsten, rapporteert onderzoek naar hoe brede maten van subjectief welzijn, zoals tevredenheid met het leven, kunnen bijdragen aan economische evaluaties in de gezondheidszorg als bron voor het schatten van de monetaire waarde van veranderingen in gezondheid. Hoofdstuk 2 maakt gebruik van gegevens uit vragenlijsten die zijn afgenomen onder de algemene bevolking van Duitsland om individuele variatie over de tijd in tevredenheid met het leven, gezondheid en inkomen te analyseren, en zo het bedrag te schatten dat nodig is om het verlies van één volledig gezond levensjaar te compenseren. Deze studie biedt de eerste schattingen van de monetaire waarde van een voor kwaliteit gecorrigeerd levensjaar voor Duitsland en behandelt verschillende methodologische aandachtspunten bij het gebruik van deze relatief nieuwe benadering.

Het tweede thema, het effect van gezondheidsveranderingen op percepties van gezondheid en gezond en gedrag, rapporteert onderzoek dat zich richt op de stabiliteit en vergelijkbaarheid van zelfgerapporteerde uitkomstmaten in de tijd, en hun relatie met objectieve informatie over gezondheid. Hoofdstuk 3 onderzoekt of zelf gerapporteerde tevredenheid met het leven en gezondheid in de tijd zich aanpassen aan het oplopen van langdurige beperkingen, gebruikmakend van een longitudinale dataset uit het Verenigd Koninkrijk. Deze studie laat een consistent patroon zien van aanpassing aan beperkingen, die vooral te zien is in de maten van breder welzijn zoals tevredenheid met het leven. Hoofdstuk 4 gebruikt een unieke combinatie van individuele gegevens uit vragenlijsten en administratieve data over gezondheid om de causale effecten te analyseren van gezondheidsschokken op percepties van gezondheid en gezond gedrag. Het toont aan dat het ervaren van een hartaanval of beroerte leidt tot blijvende veranderingen in perceptie van gezondheid en tot toename van preventief gedrag door het verminderen van risicovolle gedragingen zoals alcoholgebruik en roken.

Het derde thema, de indirecte effecten van afnemende gezondheid, rapporteert onderzoek naar de gevolgen van gezondheidsveranderingen voor directe familieleden. De tweede helft van Hoofdstuk 4 bestudeert de impact van hartaanvallen en beroertes op de percepties van gezondheid en gezond gedrag van.partners van patiënten In tegenstelling tot de significante gedragsaanpassingen bij patiënten, vindt deze studie slechts beperkt bewijs voor vergelijkbare aanpassingen bij de partners van deze patiëntengroepen. Gezien de implicaties van gezondheidsschokken voor vraag naar en aanbod van zorg binnen en tussen huishoudens, richten Hoofdstukken 5 en 6 zich op de rol van mantelzorg en de impact ervan op de gezondheid van mantelzorgers, als indirecte gevolgen van afnemende gezondheid. Met behulp van gegevens uit vragenlijsten uit het VK en Nederland onderzoeken deze hoofdstukken de korte- en langetermijn effecten van mantelzorg op de gezondheid van mantelzorgers, en toont causaal bewijs voor aanzienlijke en aanhoudende verliezen in mentale gezondheid bij mantelzorgers die intensieve zorg verlenen.

De bevindingen uit de hoofdstukken van dit proefschrift tonen de multidimensionale impact van gezondheidsveranderingen aan en laten zien hoe subjectieve uitkomstmaten uit vragenlijsten kunnen helpen deze gevolgen te begrijpen en gezondheidsbeleid te informeren. Als maatstaf van ervaren waarde kunnen ze bijdragen aan het schatten van de monetaire waarde van gezondheidsveranderingen (Hoofdstuk 2) en hoe mensen zich met de tijd aanpassen aan gezondheidsveranderingen (Hoofdstuk 3). Tegelijkertijd kunnen maten van gepercipieerde gezondheid gecombineerd met gedetailleerde informatie over objectieve gezondheidsveranderingen, verduidelijken hoe individuele besluitvorming wordt beïnvloed door gezondheidsschokken, en de mate waarin blijvende leer- en gedragseffecten optreden (Hoofdstuk 4). Tot slot, kunnen subjectieve uitkomstmaten uit vragenlijsten ook helpen bij het identificeren van de omvang van indirecte effecten op gebieden die moeilijk te kwantificeren zijn of onder populaties die gewoonlijk niet worden meegenomen in routinematig verzamelde administratieve gegevens, door een schatting te bieden van de effecten van het verlenen van informele zorg op fysieke en mentale gezondheid onder verschillende groepen mantelzorgers (Hoofdstukken 5 en 6).

About the Author

Jannis Stöckel was born in Quakenbrück, Germany, in 1993. He started his academic journey at the University of Mannheim, where he pursued a Bachelor of Science degree in Economics, graduating in 2015. During undergraduate studies he spent four months at the Erasmus School of Economics in Rotterdam, the Netherlands, and worked as a part-time research assistant at an interdisciplinary research centre at the University of Mannheim. Subsequently he completed a Master of Science in Economics in 2017 and a Master of Arts in European Studies in 2018 at the KU Leuven in Belgium. During this period, he worked as a research assistant at the Faculty of Economics and Business. Additionally, he was actively involved in Academics for Development Leuven, a student-run, non-profit consultancy aiding NGOs and businesses in developing countries, culminating in his service as Vice-President/Project Coordinator during the academic year of 2016/2017.

In January 2018, Jannis returned to Rotterdam to begin his doctoral studies at the Erasmus School of Health Policy & Management as one of multiple PhD students contributing to the multidisciplinary Erasmus Initiative Smarter Choices for Better Health. His research interests during this time can be broadly defined as understanding the impact of health changes over the life course on individual-level well-being and the broader spillover effects health shocks can create within social networks. The research conducted during this period has contributed to the chapters of this dissertation. During his time as a PhD candidate, he served as the chairman of Young ESHPM and spent two terms at the London School of Economics as a visiting research student in Health Economics and Health Policy.

Since April 2022, Jannis is working as a Research Officer at LSE Health, a multidisciplinary research centre at the Department of Health Policy at the London School of Economics. He maintains close ties with the Erasmus School of Health Policy & Management through his ongoing collaborations with his former colleagues.

Portfolio

Publications	Year
Bom, J., & Stöckel, J. (2021). Is the grass greener on the other side? The health impact of providing informal care in the UK and the Netherlands. *Social Science & Medicine*, 269, 113562.	2021
Himmler, S., Stöckel, J., van Exel, J., & Brouwer, W. B. (2021). The value of health—Empirical issues when estimating the monetary value of a quality-adjusted life year based on well-being data. *Health Economics*, 30(8), 1849-1870.	2021
Stöckel, J., & Bom, J. (2022). Revisiting longer-term health effects of informal caregiving: Evidence from the UK. *The Journal of the Economics of Ageing*, 21, 100343.	2022
Stöckel, J., van Exel, J., & Brouwer, W. B. (2023). Adaptation in life satisfaction and self-assessed health to disability-Evidence from the UK. *Social Science & Medicine*, 328, 115996.	2023

Training Activities and Courses	
PhD Basic Didactics *RISBO*	2018
Measurement of Patient Preferences using Discrete Choice Experiments *Erasmus School of Health Policy & Management*	2018
Panel Data Econometrics: Theory and Practice (Summer School) *Erasmus Research Institute of Management*	2018
How to Manage your PhD Project *Erasmus Graduate School of Social Sciences and the Humanities*	2018
English Academic Writing *Erasmus Graduate School of Social Sciences and the Humanities*	2019
Winter School in Data Analytics and Machine Learning *University of Fribourg*	2021
Data visualization – the Art/Skill Cocktail *Max Planck Institute for Demographic Research*	2021
Digital Research Methods for Textual Data *Erasmus Graduate School of Social Sciences and the Humanities*	2021

Teaching Activities		
Measuring Patient Preferences using Discrete Choice Experiments *Erasmus School of Health Policy & Management*		2019
Advanced Research Methods *Erasmus School of Health Policy & Management*		2019
Economics of Health and Healthcare *Erasmus School of Health Policy & Management*		2019 to 2022
Public Health Economics *Erasmus School of Health Policy & Management*		2020 to 2022
Master Thesis Supervision *Erasmus School of Health Policy & Management* *Erasmus School of Economics*		2019 to 2022
- Health Economics, Policy, and Law	(9/10 graduated)	2019 to 2022
- Health Economics	(2/2 graduated)	2022
- Health Sciences, Economics Track	(1/1 graduated)	2019

Presentations at Conferences and Workshops

11th Annual Meeting of the German Health Economics Association	2019
11th lowlands Health Economists' Study Group	2019
UK Health Economists' Study Group Summer Meeting 2019	2019
15th Congress of the International Health Economics Association	2019
European Health Economics Association PhD/Early Career Researcher Workshop	2019
UK Health Economists' Study Group Winter Meeting 2020	2020
NETSPAR International Pension Workshop (invited discussant)	2020
16th Congress of the International Health Economics Association (online, invited discussant)	2021
European Health Economics Association PhD/Early Career Researcher Workshop (online)	2021
14th Annual Meeting of the German Health Economics Association	2022
CEPRA/NBER Workshop on Ageing, Cognitive Ability, and Decision Making	2022
11th Annual Conference of the American Society of Health Economists	2022
2nd Health Economics Workshop, University of Copenhagen	2022

Presentations at Seminars

KU Leuven, Department of Economics (Public/Labor/Health Economics Seminar)	2019
London School of Economics, Care Policy Evaluation Centre	2020
European Health Economics Association, Spring Seminar Series	2022
University of Manchester, Health organisation, policy, and economics Seminar	2022
University of Hamburg, Hamburg Centre for Health Economics Research Seminar	2022
London Health Economics Study Group	2023
University of Copenhagen, Department of Economics (CopHEN Seminar)	2023

Journal Reviewing

Health Economics (2)	2021, 2022
The Journal of the Economics of Ageing (2)	2021, 2023
Labour Economics	2021
Journal of Economic Behavior & Organization	2022
Social Science & Medicine (2)	2022, 2023

Grants, Fellowships and Awards

RAND Summer Institute International Fellowship (*deferred, participation pending*)	2020
Erasmus Trustfonds Research Visit Grant	2021
Convergence Health and Technology Open Research Award (*joint with Sebastian Himmler*)	2021
Smarter Choices for Better Health Seed Grant	2022

Other Activities

young ESHPM, Chairman of the Board	2020
Initiator/Organiser of the ESE/ESHPM PhD Symposium (*joint with Joaquim Vidiella-Martin*)	2019, 2020

Acknowledgements

The cover of this dissertation is a nod to the painting *The Treachery of Images* by surrealist Belgian painter René Magritte that most people know with famous scription "*Ceci n'est pas une pipe*" (this is not a pipe) underneath. Its central message is simple: The painting of an object is not the object itself. This is not only good summary of the core limitation applicable to any chapter in this dissertation, relying on survey data to approximate real-life "objects" of interest, but it also has relevance for how I look back at my PhD: The included chapters, luckily, are only a (partial) representation of how my past five years have been spent, and what made it worth doing. Filling this gap is what an acknowledgements section is for so I want to use these last pages to say thank you to those on whose support I could rely along the way.

I first want to thank my supervisors, Job, Werner, and Esther, without whom there would be no dissertation. Esther, while you were more involved in the initial phase of my PhD diverging research interests resulted in us to increasingly walking different paths. Nonetheless, I cannot overstate the impact you had on me as I was taking my first steps at ESHPM. Despite never formally working together you remained a generous source of advice on all aspects of what it means to be a researcher. Without your encouragement and support I do not think I would have seen the end of it. Werner, as any PhD student who had the pleasure of having you as a supervisor can attest it is hard to not enjoy doing this research thing when having frequent interactions with you. The speed at which you effortlessly switch between lightening the mood after a setback to chart a constructive way forward speaks to both your depth of knowledge, pragmatism, and genuine belief that despite all frustrations one should not lose sight of the important things, and never take oneself too seriously. Lastly, I need to thank Job. When coming to Rotterdam before officially taking up my position in November 2017 you were the first person I met and like most socialised in Germany I had just the right mix of slight anxiety about authority and respect when talking to a "Herr Professor Dr" about wanting to do a PhD with him. Luckily, meeting you makes it immediately clear that while the title fits you well you are no person for credentialism. Throughout the past years I could always count on you for honest advice and never hesitated to seek it knowing that no matter how full your calendar was you would always find time for any of your PhDs seeking council. All three of you together I want to thank for generally believing in me and giving me not only the space but also the confidence and support necessary to chart my own path forward. Irrespective of your direct involvement in any research project your feedback was always generous and helpful, and you never ceased to point me in the right direction or opened opportunities. Thank you for this.

Next to my supervisors there are others that also contributed to the chapters in this dissertation and with whom I had the privilege to work with. Sebastian, despite you being from the wrong side of the line between Aldi-North and -South Germanys we not only managed to arrive at a fruitful research collaboration, but also had countless Saturdays spend together doing some work and/or sport and rewarding ourselves with Gözleme on the weekly market. Without your pragmatic and goal-oriented approach I am sure it would have mainly been rewards with little

justification for it. In a similar vein I need to thank Judith, with whom I had the pleasure to write two chapters in this dissertation. You taught me a lot when it comes to being a little more efficient and a little less obsessive and I got to know you not only as an insightful and generous person but also as somebody with whom one can always share a good laugh. Lastly, I want to thank Pieter and Bram. Where Judith and Sebastian brought out the best in me in terms of completing projects you made the mistake of introducing me to the Dutch administrative data and its sheer endless possibilities to tinker. I hope I have yet to reach the limits of your patience as working with you both has been a pleasure and a rewarding learning experience. Not only are both of you great economists but also generous colleagues on whose office door I would not hesitate to knock when looking for advice, whether career or research related, if it was not always open in the first place.

Apart from those that I directly worked with there is also a range of people that have made my time on campus infinitely more enjoyable. First is Stefan. Despite me not returning the honour of making you one of my paranymphs you offered to help with organising my defense without batting an eye. Colleagues who remember a nasty new scar on my nose at some point might associate this generosity in part to bottled-up guilt after my face had a run-in with your wide-swinging squash-play, but I would contest this. Thank you for being such a good and generous friend and colleague throughout the past years. While much of my time was spent at ESHPM I was lucky to be able to rely on Joaquim who kept me tethered to the health economists over at the ESE. The distance Oxford-London is not the same as a walk across campus but it’s good to have somebody on the Island nearby to occasionally share memories of the golden times of doing a PhD in Rotterdam where finally we do not have to pretend to run a two-person journal reading club anymore to justify grabbing a pint. Besides these two there are many other colleagues to say thank you to; Marianne, for playing the lightning-rod for many a question around research I fired at you; Sebastian, the coffee-one, for always being available for a chat over your favorite beverage; Marlies and Arthur, for being great office mates who together with me made for just the right mix of chattiness one would hope for; Pieter van Baal, for the occasional visit to steal sweats from the common stash and dish out some of your wisdoms; Klas, for being an inspiring stoic; Meg, for adding just the right amount of British humour and keeping the social events running; and all others who I now do not name directly through no fault of your own but my own failing memory.

There were also many friends beyond Woudestein that made my time exploring Rotterdam more enjoyable but most important were my former flat mates Xavi and Vidushi. When moving to Rotterdam I was faced with the same dreadful prospects that until today make “Rotterdam rental offerings” a cursed search on Facebook marketplace. Fortunately, when Xavi looked for new flat mates, he simply stopped interviewing after having met two that were seemingly sane enough. Vidushi joined from Hawaii and I from Germany with little knowledge of the new place or our new roommates apart from a grainy video-call. This chance encounter resulted in us living together for multiple years and that should be enough said. Whether binge watching series, rolling out the yoga mats, or just plain old doing nothing, I could not think of a better place than SMK 127 for it. Living like a student for a few more years but with the added benefit of having money and not being forced to constantly talk about what happens at university as your flat mates work real jobs is one of the greatest perks of doing a PhD. Thank you for all the good memories.

Outside of Rotterdam I could also rely on many friends with whom contact might have become less frequent but never less rewarding. I want to thank Cam for our time together in Leuven before we both started our PhDs and our travels thereafter. Road-tripping across New Zealand in two weeks and the entire US West-Coast in three has been exhausting but something I will remember forever, and thanks to our discussions one could even count it as research time (of sorts). Hela and Fabian I need to thank for not only being around since our time in Mannheim but also for agreeing to play an integral role during my defense as my paranymphs. Thank you, Fabian, for being a friend I can always rely on, for our workcation in Scotland and for making my first year living in London as pleasant as it was; you are sorely missed over here. Thank you, Hela, for being a friend I can always confide in and while we both got a bit older since our common coffee breaks at the Econ and Law library in Mannheim and distance sometimes makes keeping in-the-know on day-to-day things a bit harder I know that no matter what I can always count on you. And besides you two, thank you to all others who came along with you to join me in Rotterdam on this special day.

Natürlich muss ich auch meiner Familie danken, die mir mit viel Geduld zur Seite stand. Am wichtigsten ist da meine Mutter. Mein Erfolg wäre nicht möglich gewesen ohne dich. Du hast mich nicht nur gut erzogen (hoffe ich), sondern schlicht viel geopfert damit ich heute stehen kann, wo ich bin, und den Weg gehen konnte, den ich wollte. Und an deiner, und meiner Seite steht seit langer Zeit Heinz-Hermann, dem ich genauso danken will für seine Unterstützung über all die Jahre. Seitdem ich Fürstenau verlassen habe, hat sich mit jeder neuen Universität der Reiseweg nicht immer geografisch, aber sicher gefühlt etwas verlängert. Selbst wenn die nächste Station wieder nicht in Deutschland ist, weiß ich das egal wie weit ich weg bin, wie beim Drachensteigen bleibt ihr beiden die Schnur, die mich fest am Boden hält. Danke für Alles.

Lastly, I need to thank Nola. One of your first reactions to me starting a PhD was to google "*tips for dating a PhD student*", no doubt with some anxiety given your parents are both former academics with their own stories to share. This made you already more prepared than me, and in no small part contributed to both of us seeing the end of this together. Every presentation I held you heard before (more than once), and every swing of the pendulum between a state of elevation after a presumed discovery and the inevitable dopamine drop you managed with grace, compassion and most importantly, patience. Since our brief time in Leuven the flights Warsaw-Amsterdam turned into trains from and to London and last year finally into a pair of keys for our personal Number 10 to start our little London life together. Words cannot describe how happy I am about having you as my partner. Thank you for bringing out the best in me.

Jannis Stöckel, London, November 26th, 2023