



Stowasser, Till; Heiss, Florian; McFadden, Daniel und Winter, Joachim:

"Healthy, wealthy, and wise?" revisited: An analysis of the causal pathways from socio-economic status to health

Munich Discussion Paper No. 2014-15

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at <https://doi.org/10.5282/ubm/epub.20846>

“Healthy, wealthy, and wise?” revisited

An analysis of the causal pathways from socio-economic status to health

Till Stowasser*, Florian Heiss[†], Daniel McFadden[‡] and Joachim Winter[§]

May 14, 2014

Abstract

Much has been said about the stylized fact that the economically successful are not only wealthier but also healthier than the less affluent. There is little doubt about the existence of this socio-economic gradient in health, but there remains a vivid debate about its source. In this paper, we review the methodological challenges involved in testing the causal relationships between socio-economic status and health. We describe the approach of testing for the absence of causal channels developed by Adams et al. (2003) that seeks identification without the need to isolate exogenous variation in economic variables, and we repeat their analysis using the full range of data that have become available in the Health and Retirement Study since, both in terms of observations years and age ranges covered. This analysis shows that causal inference critically depends on which time periods are used for estimation. Using the information of longer panels has the greatest effect on results. We find that SES causality cannot be ruled out for a larger number of health conditions than in the original study. An approach based on a reduced-form interpretation of causality thus is not very informative, at least as long as the confounding influence of hidden common factors is not fully controlled.

Keywords: health; wealth; socio-economic gradient; causal inference; Granger causality; individual heterogeneity.

JEL classification: C33; I0; I12.

Acknowledgements: We are grateful to Bob Willis and participants of the 2011 NBER Aging Conference for helpful comments. Financial support was provided by the National Institute on Aging (NIA) grant No. P01 AG005842.

*Ludwig-Maximilians-Universität München. E-mail: till.stowasser@econ.lmu.de.

[†]Johannes-Gutenberg-Universität Mainz

[‡]University of California, Berkeley and University of Southern California

[§]Ludwig-Maximilians-Universität München

1 Introduction

In health economics, there is little dispute that the socio-economic status (SES) of individuals is positively correlated with their health status. The size of the body of literature documenting that wealthy and well-educated people generally enjoy better health and longer life is impressive.¹ The robustness of this association is underscored by the fact that the so-called health-wealth gradient has been detected in different times, countries, populations, age-structures, and for both men and women. Moreover, the results are largely insensitive to the choice of SES measures (such as wealth, income, education, occupation, or social class) and health outcomes.

While the existence of the gradient may be uncontroversial, the same cannot be said about its explanation. Medical researchers, economists and other social scientists have developed a large number of competing theories that can broadly be categorized as follows: there may be causal effects from SES to health, causal effects that work in the opposite direction, and unobserved common factors that influence both variables in the same direction without a causal link between the two. Distinguishing among these explanations is important since they have different implications for public policy aimed at improving overall well-being. For instance, if causal links between wealth and health were confirmed, society would likely benefit from more universal access to health care and redistributive economic policy. Yet, if such causal links were rebutted, resources would be better spent on influencing health knowledge, preferences, and ultimately the behavior of individuals.

Besides its importance, the discrimination between these alternative hypotheses also poses a great methodological challenge since the variation found in observational data is typically endogenous. This is especially true for cross-sectional data, which only offers a snapshot of the association between health and wealth. Without further information on the history of both variables, the researcher faces a fundamental simultaneity problem, which makes the identification of causal paths a hopeless venture. A possible remedy consists of finding some sort of exogenous variation in SES or health to infer causality and the direction of its flow. This search, however, is typically quite difficult because convincing instrumental variables are very hard to come by. As a consequence, researchers often face the unattractive choice between the easy path of ignoring the endogeneity problem, which casts serious doubts on any drawn conclusions, and the more involved use of IV strategies that critically rely on the untestable quality of the instruments.

¹? and ? provide extensive surveys of the earlier literature. A brief summary of more recent contributions to this field can be found in ?.

The nexus of health, wealth, and wisdom is also the subject of the study by ? (HWW henceforth). The authors propose an innovative approach that attempts to solve the above trade-off, on the premise that causal inference may be possible without having to isolate exogenous variation in SES. Their identification strategy consists of two main ingredients: First, they exploit the dynamic nature of panel data, focusing on health *innovations* rather than the prevalence of medical conditions. Second, they make use of the so-called Granger causality framework, which represents a purely statistical approach to the theory of causation. The great advantage of working with this alternative concept is that the detection of potential Granger causality is a rather easy task. While knowledge on the existence of Granger causality may not be useful in its own right, it allows for tests on the *absence* of “true” causality in a structural sense.

Applying this framework to the first three waves of the *Asset and Health Dynamics among the Oldest Old* (AHEAD) survey study, HWW find that in an elderly US population, causal channels that operate from wealth to health are an exception rather than the rule: while causality cannot be ruled out for some chronic and mental conditions for which health insurance coverage is not universal, SES is unlikely to be causal for mortality and most other illnesses. Considering these strong results, as well as the methodological novelty of HWW’s approach, it is not surprising that their work has subsequently been the subject of vivid debate within the literature.² So far, the focus has clearly been on the validity of HWW’s identification strategy in general, with some calling into question the ability to truly infer causality with a concept that arguably is a rather sparse characterization of causal properties.

We certainly agree that HWW’s model would benefit from certain methodological refinements and plan to implement these in future research. For the present project, however, we deliberately leave the econometrics unchanged, to study a different aspect that also merits attention: the stability of HWW’s results when confronted with new data that allows for hypothesis tests of greater statistical power. Special interest lies in assessing whether the somewhat surprising absence of direct causal links from SES to most medical conditions is a robust finding or perhaps the artifact of a particular data sample. Since the publication of HWW’s original article, the AHEAD survey has been incorporated into the more-encompassing *Health and Retirement Study* (HRS). This permits deviations from HWW’s data benchmark along the following dimensions: the same individuals can be tracked for a longer period of time, the analysis can be extended to new cohorts of respondents, and the working sample can be widened by including younger individuals aged 50 and older. The last point is of special interest as it offers variation in health insurance status that is not available

²As an example, consider the comments to HWW by ?, ?, ?, ?, ?, ?, ?, and ? published in the same issue as the original article.

in a Medicare-eligible population. To understand which of these data changes contribute to any deviating conclusions, we do not apply the whole bundle of modifications at once. Instead, we estimate the model multiple times, by applying it to several different data samples, which are gradually augmented along the dimensions just outlined.

We lay out the theoretical background of our analysis in section 2, where we review the potential explanations for the association between SES and health and specify the econometric challenges that arise when trying to discriminate among them. This is followed by a discussion of how to address these challenges. Section 3 describes the approach proposed by HWW. A reanalysis of HWW with new data is presented in section 4. Section 5 concludes and outlines topics for future research.

2 The difficulty of causal inference

2.1 The issue: Potential channels between SES and health

Correlation does not necessarily imply causation. This insight is one of the main lessons every empiricist needs to internalize. At times, however, it can be tempting to neglect this admonition, especially when a causal interpretation of a joint motion of two variables is very intuitive. The relationship between SES and health is a prime example for such a situation. As an illustration, consider table 1, which lists household median wealth of HRS respondents arrayed against self-reported health status. Here, the wealth-health gradient is prominently on display as median wealth monotonically decreases with impairing health self-reports – an observation that is remarkably stable over time.

What could be more natural than to interpret this strong correlation as a causal influence of wealth on health? After all, it is the explanation best in line with conventional wisdom: money can buy (almost) anything – even better health. Yet, the most intuitive conclusion may not necessarily be the only valid one. In fact, there are two additional hypotheses for the association of SES and medical conditions: the causation could flow from the latter to the former, and

Table 1. *Median wealth by self-rated health status*

Self-rated health	1992	1996	2000	2004	2008
Excellent	155.6	192.0	256.0	331.4	363.0
Very good	122.1	159.0	202.6	240.0	304.0
Good	82.5	106.2	130.6	160.0	194.0
Poor	46.7	62.2	69.0	75.1	86.1
Fair	19.5	35.0	36.5	39.7	48.1

Notes: Calculations by authors based on HRS data. Numbers reported in thousands of USD.

the correlation may actually be spurious, with third factors affecting health and wealth in a similar way. This section describes these rivalling theories and gives an overview of the most commonly-cited potential pathways between SES and health (see ?, ?, ?, and ? for more extensive reviews).

Hypothesis A: SES has a causal influence on health outcomes

This is the hypothesis most energetically advocated within the epidemiological literature. While it is true that the main contribution from economists consists of formulating alternative interpretations of the socio-economic health gradient (see hypothesis B, below), it should be emphasized that they are not on record of categorically challenging hypothesis A, either. Below, we list the most prominent theories of channels through which SES may have a causal effect on health.

Channel A1: Affordability of health care. This potential channel is arguably one of the most intuitive explanations and may be active both before and after an individual is diagnosed with an illness. For one, varying SES may be responsible for differentials in the onset of health conditions as poorer people may be overly sensitive to the costs of preventive health care. In addition, wealth could play a crucial role in determining the quality or even the plain affordability of medical treatments, once they become necessary.

Channel A2: The psychological burden of being poor. Medical scientists increasingly emphasize the importance of psychological consequences of low SES. They argue that low-wage employment is typically associated with a high degree of work monotonicity and low job control, leading to psychosocial stress. Similarly, economically disadvantaged individuals are believed to be repeatedly exposed to episodes of high emotional discomfort, either due to long phases of unemployment or a general feeling of social injustice. When accumulated, these stressful experiences may well have strong adverse effects on physical health as well. Furthermore, adverse wealth shocks – such as the loss of life savings in a stock market crash – are likely to cause anxiety and depression, representing a more immediate avenue through which SES may impact health.

Channel A3: Environmental hazards. Another line of argument is that the exposure to perilous environments is considerably higher for the poor. This may concern job-related risks since it can be argued that workplace safety is lower and physical strain higher for poorly-paid occupations. The reasoning also extends to people's living environments as neighborhood safety, dwelling condition, air and water quality, etc. are usually much better in exclusive residential areas.

Channel A4: Health knowledge. Considering that education is an integral component of SES, it is conceivable that part of the correlation between SES

and health is attributable to differences in health knowledge. According to this argument, information on medical risk factors or the importance of preventative care may be more widespread among the highly educated and wealthy, leading to healthier lifestyles and lower morbidity rates among this group.

Channel A5: Risk behaviors. An often-cited pathway through which SES may influence health is the asymmetric distribution of unhealthy lifestyles such as smoking, drinking and poor diet. To the extent that all of these vices are less common among the rich, health differentials may in fact be driven by SES variables. Note that the question of why smoking, excess alcohol consumption, and obesity are especially prevalent in lower social classes, is interesting in its own right, with channels A2 and A4 potentially accounting for part of this relationship.

Hypothesis B: Health has a causal influence on SES outcomes

Economists and other social scientists were among the first to challenge the conception that causal mechanisms would work their way exclusively from SES to health. Much of this research is inspired by ?'s (?) health production framework, which models the impact of health capital on savings, labor market participation, and retirement decisions. We believe the following three channels to be the most important in describing causal effects from health to SES outcomes.

Channel B1: Productivity and labor supply. Arguably, the most relevant reason why health may be causal for SES outcomes can be found on the labor market. The productivity of an individual in poor health is generally lower than that of someone whose physical robustness allows for longer working hours, less absenteeism, and better career options. As a consequence, frail people will tend to earn lower wages and accumulate less assets throughout their life course. Adverse health shocks may even be so severe that people are forced to leave the labor market altogether, depriving them from any realistic chance to improve their SES.

Channel B2: Life expectancy and time preferences. To the extent that severe illnesses increase mortality risks, there may be an impact of poor health on time preferences. Life-cycle models predict that the optimal response to a perceived reduction in life expectancy is to move consumption from an uncertain future towards the present. Thus, a history of dire medical events may induce individuals to dissave faster, establishing a causal link from health to SES.

Channel B3: Medical care expenditures. The most immediate form of impact health events can have on financial endowments are out-of-pocket costs of medical care. While it can be argued that the influence of this path-

way should only be modest in size, this is certainly untrue for people without health insurance. In many cases, not even the insured are completely shielded from medical bills: the existence of deductibles and lifetime coverage limits poses great financial threats especially for the chronically ill.

Hypothesis C: SES and health are jointly caused by an unobserved third factor

This hypothesis makes the case that the association between health and wealth could have other reasons than causal mechanisms between the two: There may be hidden third factors with a common influence on both SES and health, rendering the correlation among the latter spurious. This distinction is vital since policies that aim at improving health outcomes by, say, redistributing wealth are bound to be ineffective, as long as the true common cause remains unaffected.

Channel C1: Unobserved genetic heterogeneity. A good candidate for an unobserved common cause is genetic disposition. For instance, genetic frailty may reduce the physical resistance as well as the intellectual and professional skills of an individual. In such cases, health will be poorer and SES will be lower despite the absence of causal links among the two.

Channel C2: Unobserved family background. Genetic endowment is not the only determinant of people's physical and personal traits. Similarly influential are matters of parentage and upbringing. Especially pre-natal and early-childhood nutrition as well as stress are believed to have lasting negative effects on well-being and functional abilities, establishing an association between health and SES that is similar to that of Channel C1.

Channel C3: Unobserved preferences. Irrespective of whether they are inherited or learned, preferences that influence certain behavior and lifestyles are another often-cited source of common effects. The prime example are descendants of dysfunctional families, who adopt both the unhealthy lifestyles (such as poor diet or smoking) and the unambitious attitudes towards education and work by which they are surrounded. Another example are time preferences: overly myopic people will underinvest in preventative medical care and in education since in both cases pay-offs will materialize in a distant future, to which only little importance is attached.

2.2 The challenges: Simultaneity and omitted variables

The fact that all of the aforementioned hypotheses are generally plausible, makes the inference on causation a methodologically challenging task. Suppose – as is the case for the remainder of this paper – we were interested in testing the validity of hypothesis A, that is whether SES has a causal effect on health outcomes. Ideally, we would want our analysis to rely on truly exogenous variation in SES variables, similar to that attained in controlled experiments. The reality

for economists, however, is far from being ideal since the sources of variation we find in observational data is unknown to us. As a consequence, causal variables are potentially endogenous themselves.

The possible sources of endogeneity in the wealth-health case have been described in section 2.1. Ultimately, they generate two fundamental econometric challenges: we have to distinguish hypothesis A from hypothesis B, and hypothesis A from hypothesis C. As we discuss below, the first consists of dealing with a simultaneity problem, and the second of finding a solution to the problem of omitted variables.

Challenge 1: The simultaneity problem (hypothesis A vs. hypothesis B)

Imagine for a moment that hypothesis C could be dismissed, so that any association between SES and health had to be due to either hypothesis A or B. Even with this kind of simplification in place, the identification of SES causality for health is still difficult. Of course, we could regress our health variable of interest (H_i) on SES (S_i) and a vector of exogenous control variables (\mathbf{X}_i), estimating the following equation with OLS:

$$H_i = \theta_0 + \theta_s S_i + \mathbf{X}_i' \boldsymbol{\theta}_x + \eta_i, \quad (1)$$

where i denotes the unit of observation and η_i is the residual. Yet, the crucial question is if we could interpret the parameter estimate $\hat{\theta}_s$ as the causal effect of SES on health. The answer would be affirmative if the structural model were to look like

$$\begin{aligned} E(H_i | S_i, \mathbf{X}_i) &= \alpha + \beta S_i + \mathbf{X}_i' \boldsymbol{\gamma}, \\ E(S_i | H_i, \mathbf{X}_i) &= E(S_i | \mathbf{X}_i). \end{aligned}$$

This model describes a world in which causality only flows from SES to health, with β capturing the true causal effect. In this world, $\hat{\theta}_s$ would indeed have a causal interpretation, with $\text{plim } \hat{\theta}_s = \beta$. However, the existence of hypothesis B indicates that the above model may not be a realistic description of reality. In fact, the true structural model is likelier to look like

$$E(H_i | S_i, \mathbf{X}_i) = \alpha + \beta S_i + \mathbf{X}_i' \boldsymbol{\gamma}, \quad (2)$$

$$E(S_i | H_i, \mathbf{X}_i) = a + b H_i + \mathbf{X}_i' \mathbf{c}, \quad (3)$$

with β again measuring the true causal effect of SES on health, and b capturing any causation working its way in the opposite direction. Equations 2 and 3 describe a standard simultaneous-equation model (SEM) as both dependent variables are jointly determined with each being a function of the other. When trying to estimate this SEM by simply running regression equation 1, $\hat{\theta}_s$ will be subject to simultaneous-equation bias, picking up the information conveyed in

b as well. As a result, the parameter of interest, β , is not identified, making a test for causation of SES to health all but impossible.

Challenge 2: The omitted-variable problem (hypothesis A vs. hypothesis C)

Even in the absence of challenge 1, we would still face the problem of having to discriminate between hypotheses A and C. Presume we were able to plausibly exclude causal paths from health to SES. In this case, the identification problem no longer consists of confounding the causal effect of wealth on health with reverse causality. Instead, the question arises if an association between both variables is attributable to causality at all since it could also stem from a joint reaction to a third factor. As the review of hypothesis C has shown, all of these potential common causes (such as genetics or preferences) are inherently unobservable, rendering challenge 2 an omitted-variable problem.

Suppose the true structural model is best described by

$$E(H_i|S_i, \mathbf{X}_i, C_i) = \alpha + \beta S_i + \mathbf{X}_i' \gamma + \delta C_i, \quad (4)$$

with C_i standing for an individual-specific variable that influences both SES and health. If this common cause were observable, we could simply include it in our regression function and the causal effect, β , would be readily identified. However, given its omitted-variable nature, C_i will be swamped into the error term, as the comparison of the structural model in error form (equation 5) with the estimable model (equation 6) demonstrates:

$$H_i = \alpha + \beta S_i + \mathbf{X}_i' \gamma + \delta C_i + \epsilon_i, \quad (5)$$

$$H_i = \alpha + \beta S_i + \mathbf{X}_i' \gamma + u_i. \quad (6)$$

Here, the well-behaved structural error is denoted by ϵ_i , whereas the composite residual is $u_i = \delta C_i + \epsilon_i$. Given that C_i has an impact on our explanatory variable of interest, S_i , the latter will be endogenous since $\text{cov}(S_i, u_i) \neq 0$. As a consequence, the estimation of this model by means of regression equation 1 will yield a parameter estimate $\hat{\theta}_s$ that suffers from omitted-variable bias, with $\text{plim } \hat{\theta}_s \neq \beta$. Importantly, $\hat{\theta}_s$ will absorb any causal impact that C_i may have on H_i . As a result, the presence of common effects could easily lead to erroneous conclusions of active causal links between wealth and health in cases where β actually equals zero.

Causal inference in the face of both challenges

Naturally, there is no reason to believe that both econometric problems are mutually exclusive. As a rule, they will be present at the same time, aggravating

causal inference even more. Ultimately, we have to estimate a structural model that takes the following form:

$$H_i = \alpha + \beta S_i + \mathbf{X}_i' \gamma + \underbrace{\delta C_i + \epsilon_i}_{=u_i}, \quad (7)$$

$$S_i = a + b H_i + \mathbf{X}_i' c + \underbrace{d C_i + e_i}_{=v_i}, \quad (8)$$

with e_i denoting a structural error and v_i representing the composite unobservable. Given this multitude of potential confounders, we truly cannot expect the simple regression function 1 to uncover β , the structural parameter of interest. While this assessment is certainly sobering, it also sets a clearly defined bar for any alternative identification strategy: in order to be convincing, it has to live up to the challenges of simultaneity and omitted variables.

A common way of dealing with the potential endogeneity of SES is the use of instrumental variable (IV) estimators. The virtue of this approach is that – at least in theory – it solves both of these challenges at once. A good instrument is, however, hard to find in practice. In the context of the SES-health causality, exogenous wealth shocks have been used as instrumental variables. For instance, ? as well as ? use inheritances. In a similar vein, ? interprets the strong stock-market surge in the 1990s as a positive wealth shock, and it is probably just a matter of time until we will see the first papers that make use of the exogenous variation in wealth caused by the recent global financial crisis.

We do not discuss IV approaches in detail, but we would like to point out one problem that arises in the analysis of the SES-health gradient. While the above instruments may well be exogenous and certainly have an impact on wealth, it is not entirely clear if the SES variation they induce is really that relevant for *health*. According to ?'s (?) standard economic model of health, an individual's general health status can be viewed as a latent capital stock that reflects the entire history of medically relevant events and behaviors. As a result, the human body will certainly react to current influences but it will not forget how it was treated in the past either. This “memory effect” likely extends to any influence SES may have had during one's lifetime. In light of this, it is questionable whether sudden changes in wealth are really that informative when testing for causal links between SES and health. In fact, since an IV estimator makes use of exogenous variation in wealth *at one point in time* to identify β , there is a great chance that causal links from SES to health are statistically rejected even though they have been operating in the past.³ Admittedly, an IV estimator will still capture any *instantaneous* impact a wealth shock would have on health outcomes. As a renewed look at the potential causal pathways for hypothesis A suggests,

³In this light, it is not too surprising that none of the aforementioned studies using wealth shocks as an instrument for SES was able to find evidence supportive of hypothesis A.

immediate effects are most likely to arise through channel A.2 if wealth shocks are severe enough to have direct psychological consequences.

3 The approach of the HWW study

The previous section demonstrates that the identification of causal paths between health and wealth with IV approaches is not always feasible. Especially the isolation of truly exogenous and yet meaningful variation in SES poses considerable problems. On this account, HWW propose an alternative identification strategy that avoids this critical step altogether. In fact, they make use of the entire observed variation in SES variables, tacitly accepting that some of it may well be of endogenous nature. The authors argue that, in spite of this methodological simplification, their approach still allows for at least indirect inference of causal links from SES to health.⁴

Naturally, HWW need to find convincing answers to the two econometric challenges described in section 2.2. When testing hypothesis A, they face challenge 1 of excluding the possibility that any observed co-movement of wealth and health is in reality due to reverse causality. In addition, they have to tackle challenge 2 of ruling out that the association is driven by unobserved common effects.

Challenge 1: Ruling out hypothesis B

Distinguishing hypotheses A and B without the aid of instrumental variables is a difficult task. We may observe that the poor are less healthy but we have no information on which happened first: were people already poor before they got sick, or were they already sick before they became poor? With cross-sectional data that only offers a snapshot of this association, there is no way of finding out. Panel data, on the other hand, provides valuable information on transitions in health and wealth, making it possible to analyze the dynamics of their relationship and to identify the direction of the causality flow. Imagine we were to analyze the dependence of health *innovations* on past levels of SES. As long as one agrees that a cause must precede its effect, we can be sure that the (unanticipated) onset of an illness at time t cannot have caused the amount of wealth or education at time $t - 1$. Given there is any causation at work, it must flow from the past to the present, or – as in this case – from SES to health innovations.

HWW take this insight to heart by applying their framework to the first three panel waves of the aforementioned AHEAD survey study, which spans the years

⁴In their article, HWW also formulate tests on causality working in the opposite direction. However, the authors themselves are quite skeptical about this part of their analysis, admitting that it is likely subject to model misspecification. As they stop short of endorsing their own results, we follow their lead and concentrate on the more promising test of hypothesis A.

between 1993 and 1998 and is representative of the US population aged 70 and older. They propose a dynamic model of health incidence that takes the following form:

$$f(HI_{it}^j | \mathbf{H}_{it}^{k < j}, \mathbf{H}_{it-1}, \mathbf{S}_{it-1}, \mathbf{X}_{it-1}), \quad (9)$$

where i once again stands for the unit of observation (in this case: household) and the newly introduced t denotes time. The index j stands for the respective health condition as the authors apply their model to 20 different medical outcomes.⁵ The dependent variable, HI_{it}^j measures a new incidence of a given health condition.⁶ According to this model, a health innovation is potentially influenced by the following explanatory variables:

Past level of SES: The vector \mathbf{S}_{it-1} includes five SES variables, namely wealth, income, years of education, dwelling condition, and neighborhood safety. These are the variables of main interest. Conceptually, if SES had any direct causal impact on health, we would expect to observe that rich individuals are less likely to develop a new medical condition compared with poor individuals. While this finding alone would not yet prove the existence of a causal link from SES to health, confounding with reverse causality could be ruled out since \mathbf{S}_{it-1} precedes HI_{it}^j .

Past health status. New medical events are likely influenced by a respondent's health history as well. This may take the form of state dependence (e.g., past cancer influences the onset of new cancer) and co-morbidities (e.g., past cancer influences the onset of depression). For this reason, HWW control for vector \mathbf{H}_{it-1} , containing the past levels of all 20 health conditions.

Current health incidences with immediate impact. In theory, health innovations could also be influenced by contemporaneous shocks in SES or other health conditions. This constitutes a problem for HWW's concept of dealing with simultaneity as it critically relies on the ability to observe the timing of innovations in both variables. HWW solve this problem by imposing further structure: On the one hand they make the assumption of no instantaneous causation of SES to health shocks, arguing that any causal action as described by channels A.1 to A.5 takes time.⁷ On the other hand, they impose a chain structure on contemporaneous health innovations, grouping

⁵These include acute illnesses (cancer, heart disease, stroke), mortality, chronic conditions (lung disease, diabetes, high blood pressure, arthritis), accident-related events (incontinence, severe fall, hip fracture), mental problems (cognitive impairment, psychiatric disease, depression), as well as information on interview status (self vs. by proxy), BMI, smoking behavior, ADL/IADL impairments, and self-rated health.

⁶Note that this measure of health innovation *cannot* be interpreted as a simple change in health status ($\Delta H_{it}^j = H_{it}^j - H_{it-1}^j$) since HI_{it}^j generally captures deteriorations in health only. For chronic illnesses, such as diabetes, it measures when the condition was first diagnosed. For acute health events, such as stroke, HI_{it}^j indicates every new occurrence.

⁷The authors themselves make the point that this assumption loses its innocuousness if the time intervals between panel waves become too large since even the more inertial causal links

them in the order in which instantaneous causality is most likely to flow.⁸ Thus, they include the vector $\mathbf{HI}_{it}^{k < j}$ containing the incidence variables for all health conditions $(1, \dots, k)$ that are causally arranged upstream of condition j .

Demographic control variables. Finally, the authors control for a number of demographic factors that could have an impact on health events, too. The corresponding vector, \mathbf{X}_{it-1} , includes the respondent’s age, marital status, as well as information on the parent’s mortality and age at death.

Building on model 9, HWW design a test for non-causality of SES in the spirit of ? and ?. This so-called Granger causality (or G-causality) approach is a purely statistical take on the concept of causation, having its origin in the time-series literature. Formally, SES is not Granger causal for health condition j if

$$f(HI_{it}^j | \mathbf{HI}_{it}^{k < j}, \mathbf{H}_{it-1}, \mathbf{S}_{it-1}, \mathbf{X}_{it-1}) = f(HI_{it}^j | \mathbf{HI}_{it}^{k < j}, \mathbf{H}_{it-1}, \mathbf{X}_{it-1}), \quad (10)$$

i.e., HI_{it}^j is conditionally independent of \mathbf{S}_{it-1} , given $\mathbf{HI}_{it}^{k < j}$, \mathbf{H}_{it-1} , and \mathbf{X}_{it-1} . Intuitively, given health history, knowledge of SES history must not contribute to the predictability of health innovations. The test is implemented by estimating the model by maximum likelihood (ML) both unconstrained (with \mathbf{S}_{it-1} as regressors) and constrained (without \mathbf{S}_{it-1}) and by subsequently comparing the log likelihoods of both versions. The motivation for this likelihood ratio test is that the two values should be the same if the null hypothesis of non-causality is true.

The detection of Granger causality, however, does not guarantee the presence of “true” causality in a structural sense, which is the concept we are ultimately interested in.⁹ Admittedly, information on the presence of G-causality is helpful when predicting health innovations for an individual with *given* health and SES history. However, the reduced-form nature of G-causality renders it unsuitable to predict the effects of (economic) policy interventions. If SES is Granger causal for health innovations, we only know that, for instance, the onset of an illness is likelier for a person with low SES. Yet, we do not know if

will then have enough time to unfold. Given that the AHEAD study is conducted biennially, the time aggregation to observation intervals may indeed reintroduce some degree of simultaneity.

⁸HWW list cancer, heart disease, and stroke first because they can have an immediate impact on mortality. The other medical conditions are grouped such that degenerative illnesses can cause chronic diseases, which in turn may influence accidents and finally mental health. Importantly, instantaneous causality is not designed to flow in the opposite direction.

⁹There are three major “schools” of causal analysis: The structural approach (S-causality) described by ? and ? that is grounded in econometric simultaneous equations models, the potential-outcomes approach (P-causality) characterized by ? and ? that is based on the analysis of experimental treatments and the time-series prediction approach (G-causality) employed here. The conventional interpretation of “true” causality is arguably best described by S- and P-causality treatments. In fact, ? demonstrates a formal equivalence between the two concepts. Both of these schools are critical of G-causality, arguing that its purely positivistic approach does not realistically characterize causal properties.

this statistical dependence is due to a real causal link from wealth to health (hypothesis A) or due to unobserved common effects (hypothesis C). Given the diverging policy conclusions both interpretations would trigger, HWW also need to address the second methodological challenge of dealing with the omitted-variable problem.

Challenge 2: Ruling out hypothesis C

Most of the omitted variables identified in section 2.1 to potentially have a common influence on health and SES are unobservable by definition. As a result, challenge 2 cannot simply be resolved by improvements in data quality and the addition of missing variables to the vector of covariates. HWW also refrain from making use of fixed-effects estimation, which represents another common strategy to heal omitted-variable bias in cases where panel data is available. In fact, the efforts made by the authors to distinguish between structural causality and common effects are limited to using a rich set of covariates in the hope that this will mitigate the importance of unobservables. They argue that,

[f]or example, genetic frailty that is causal to both health problems and low wages, leading to low wealth, may be expressed through a health condition such as diabetes. Then, onset of new health conditions that are also linked to genetic frailty may be only weakly associated with low wealth, once diabetic condition has been entered as a covariate.

Despite this conciliating argument, HWW acknowledge that the failure to cleanly identify causal structures questions their approach's ability to gauge the effects of "out-of-sample" policy changes. To address this issue, they scrutinize the generality of their results by adding invariance tests to the analysis. Intuitively, a model is only suitable for the sort of predictions HWW have in mind if it remains valid under different scenarios than those covered by the data, or – as the authors put it – if it has the invariance property of being valid for each possible history. For instance, if the application of the model to different populations, time periods, and policy regimes had a negligible impact on estimation results, there would be reasonable hope that the Granger non-causality tests are indeed informative. The invariance tests as implemented by HWW mainly inspect the stability of findings across time. Model 9 is estimated by stacking the data for the two available panel wave transitions (i.e., $W1 \rightarrow W2$ and $W2 \rightarrow W3$) above another. The same model is also estimated for both wave transitions individually, and a test statistic is constructed that compares the log-likelihoods of these three estimations. The motivation for this likelihood ratio test is similar to that of a Chow test. If the null hypothesis of model invariance is true, estimated parameters of the stacked model should not differ from those of the two single-transition models.

All told, HWW apply the following system of non-causality and invariance tests to the estimations of all 20 health conditions: First, they test for Granger non-causality of SES for health innovations in the stacked version of the model under the maintained assumption of invariance (S|I). Then, they employ an unconditional invariance test, as described above (I), followed by an invariance test with non-causality imposed (I|noS). Finally they implement a joint test of invariance and non-causality (S&I). Conceptually, HWW condition the validity of their non-causality tests on the outcome of the corresponding invariance test: only if invariance is confirmed, they will put faith in the model's results. The authors are optimistic that with these refinements in place, their model is well-placed to make meaningful predictions even if it fails to identify true causal links, stating that

[i]t is unnecessary for this policy purpose to answer the question of whether the analysis has uncovered a causal structure in any deeper sense. Econometric analysis is better matched to the modest task of testing invariance and non-causality in limited domains than to the grander enterprise of discovering universal causal laws. However, our emphasis on invariance properties of the model, and on tests for Granger causality within invariant families, is consistent with the view of philosophers of science that causality is embedded in "laws" whose validity as a description of the true data generation process is characterized by their invariance properties.

They even go a step further and suggest that their approach – while not powerful enough to distinguish between causation and common effects – permits at least the one-sided test for the *absence* of true causal links. Essentially, they view Granger causality as a necessary but insufficient condition for a structural causal pathway from SES to health. Their decision criteria when interpreting results are as follows: If the invariance test fails, one should question the validity of the model for this particular health variable and refrain from drawing any conclusions. If invariance holds and Granger causality is present, one cannot distinguish between a direct causal link and a common factor. Yet, if invariance holds and Granger causality is ruled out, it should be safe to deduce that SES does not have a causal impact on the health condition under consideration.

Summary of HWW's findings

Contrary to conventional wisdom, the evidence from applying HWW's approach to the elderly US population is not *universally* supportive of hypothesis A. In fact, they find that SES is unlikely to be causal for mortality, most acute health conditions, accidents, and a large number of degenerative diseases. Medical conditions, for which direct causal links cannot be ruled out, include self-rated health status, most mental illnesses and some chronic conditions such as diabetes, lung disease and arthritis. This pattern loses some of its mysteriousness

when viewed in the context of US health-policy characteristics. The population under examination is of advanced age and eligible for Medicare, which will likely weaken any causal impact wealth could have on well-being via the affordability of health care. Yet, even Medicare coverage is not fully comprehensive and tends to focus on acute care procedures, while generally failing to limit out-of-pocket costs for treatments of chronic and psychological conditions.¹⁰ This lends indirect evidence for the importance of channel A.1 since the socio-economic gradient emerges exactly for those health conditions, for which the ability to pay is most likely to be an issue.

Reflecting the substantial degree of ambiguity in these results, the policy conclusions formulated by HWW are rather contained in both phrasing and substance. On the one hand, they cannot overcome the methodological challenge of inferring true causality when G-causality is detected. This leaves open whether SES-linked preventive care induces onset of chronic and mental illnesses or whether persistent unobserved factors are to blame for the observed health-wealth association. On the other hand, even convincing evidence for the absence of direct causal links might not necessarily warrant the bluntest form of policy recommendation. Sure enough, SES-linked therapies for acute diseases do not appear to induce health and mortality differentials, which – to quote HWW – should *theoretically* permit the strong conclusion that

policy interventions in the Medicare system to increase access or reduce out-of-pocket medical expenses will not alter the conditional probabilities of new health events[.]

However, the authors stop short of actually drawing this conclusion, which reflects their reluctance to base overly aggressive policy proposals on a concept whose ability to simulate the effect of system shocks is not indisputable.

Discussion of HWW's approach

All things considered, what should we make of HWW's approach of inferring causality and yet avoiding the cumbersome search for exogenous variation in SES? Does their reliance on Granger causality and their decision to focus on health innovations really do the trick of solving the endogeneity problem, or have they entered a methodological dead-end street? Overall, the response within the literature has been fairly critical, albeit not excoriating, pointing out a number of issues briefly discussed below.

Existence vs. activation of channels It is important to understand the limitations of an approach that focusses on *innovations* in health, rather than health status itself. HWW detect a strong and ubiquitous association of SES and prevalence of health conditions in the initial wave of their sample.

¹⁰Note that the study was conducted well before the introduction of Medicare Part D in 2006 that especially benefited the chronically ill by improving the coverage of prescription drugs.

This suggests that the elderly population under consideration has potentially been affected by some of the causal channels between health and wealth in the past. This history, however, remains a blind spot for HWW's model: by concentrating on future health events, they are unable to explain what factors lead to the pre-existing SES gradient. By contrast, they study the question whether SES has an impact on the onset of *additional* medical conditions, given an individual is already old, still alive, and has gone through a long and unexplained health-wealth history. While the analysis of an elderly population is not illegitimate and certainly interesting in its own right, one should entertain some doubts about its external validity. In theory, HWW's findings could – if extrapolated backwards – also provide a retrospective explanation for the early relation between SES and health. However, as pointed out by ?, ?, ?, and HWW themselves, this extreme form of time invariance over the entire life cycle is unlikely to hold as certain causal channels are probably relevant at different stages in one's life.¹¹ In light of this, an accepted non-causality test should perhaps not be taken as evidence against the plain existence of a causal link but rather against its *activation* within the class of invariances under consideration.

Unobserved common effects. As argued above, the major weakness of HWW's approach is that it cannot separate true causality from hidden common effects. Yet, according to the authors, this will only constitute a problem if Granger causality is detected. In the absence of G-causality, causation in a structural sense should be ruled out as well. This interpretation implies that the detection of conditional dependence is a prerequisite for an active causal link – an assumption that is questioned by ?, who argues that persistent hidden factors may also work in the opposite direction of causal pathways and offset them. If this were the case, information on G-causality might actually not tell us anything about true causal mechanisms, rendering HWW's strategy ineffective. However, the likelihood of direct causal effects being *exactly* offset by unobserved common factors should be practically zero, making this argument irrelevant for identification. Then again, there are obvious limits to this defence in finite samples, so that statistical inference of causation could indeed be seriously jeopardized by the failure to account for hidden common causes.

Invariance tests. Anticipating that their framework might fall short of inferring deep causal structures, HWW subject their model to the aforementioned invariance tests. On a conceptual level, model invariance would

¹¹For retirees, pension income is not affected by (contemporary) ability to work, occupational hazards vanished on the day of retirement, and Medicare provides basic health insurance, rendering channels B.1, A.3, and A.1, respectively, of little importance when late in the life cycle. At younger ages, however, all of these pathways may well have played an important role.

arguably justify predictions of policy effects but there are legitimate concerns whether the actual tests implemented in their paper are statistically powerful enough. ?, ?, ? and, once more, HWW themselves point out that invariance under historical interventions is of little use when the panel is as short as AHEAD, offering hardly any in-sample variation in populations, age structures, and – most importantly – policy regimes. As a consequence, an accepted invariance test as implemented by HWW is unlikely to be a sufficient condition for the sort of model validity necessary to make out-of-sample predictions. On top of that, ? even questions whether one should view the acceptance of HWW’s invariance tests as a *necessary* condition for meaningful analysis. Instead of discarding results when invariance tests are rejected, one could follow up on the reasons for time invariance failures as they may be informative of structural breaks in causal relationships. For instance, certain causal pathways may switch on or off in the course of policy changes or as the observed cohort grows older. In such cases, failed invariance tests would actually shed light on the circumstances under which causal links will be active or unexpressed, allowing for sharper, channel-specific causality tests.

Health dynamics. Another reason for concern is the fact that HWW model health dynamics as a first-order Markov process, which cannot be expected to properly capture the medium and long-run evolution of health. Intuitively, this is because the Markov model assumes that all relevant information about the whole past is captured in the observed variables one period ago. This is unrealistic since knowledge of longer histories would better capture the stock characteristics of health capital as envisioned by ?. Taking functional limitations as an example, a respondent who reported difficulties with walking one year ago and no limitations previously has a different outlook than a respondent who consistently reported difficulties with walking for the last ten years.

Instantaneous causality. Finally, ?, ?, and ? express their skepticism about HWW’s handling of instantaneous causality. The hierarchy imposed on health conditions (with the assumption that incidence of each condition is conditioned on upstream incidences but not on downstream ones) may be acceptable as a reduced-form assumption and is etiologically fairly reasonable. Yet, it likely falls short of the structural stability explored by invariance tests and is a potential source of serious model misspecification, making it a prime target for methodological improvements in the course of future research.

4 Reanalysis of HWW with new data

The preceding discussion indicates that HWW’s approach of disentangling the association between health and wealth while avoiding the often futile struggle of finding exogenous variation in SES comes at the price of limited methodological persuasiveness. However, since the generic alternative – instrumental variables – is not exempt from substantial criticism either, we certainly feel that this identification concept merits methodological refinement rather than being dismissed altogether. Some weaknesses, such as the treatment of common effects, health dynamics, or instantaneous causality, require significant modifications to the original model and we plan to implement these in future research.

Yet, one of the major downsides of HWW’s study – the lack of invariance test power – can be addressed without the need for complex changes but instead by applying the largely unaltered model to a more apposite set of data. Recall that the root of this problem is that the invariance tests are based on rather limited variation in “histories” of states relative to the universe of potential histories. Increasing the N as well as the T dimension of the panel data will arguably raise the number of histories and enhance the power of these tests. Of course, we can also expect larger sample sizes to boost the statistical power of non-causality tests, effectively reducing the risk of committing type-II errors. But sample size is not everything. We believe that the analysis will also greatly benefit from larger sample “diversity”, with data covering different kinds of populations that are subject to varying institutional setups. For instance, the inclusion of younger respondents could shed light on the question if the activation of causal links is stable throughout the life cycle or if reaching the retirement age induces some sort of structural break.

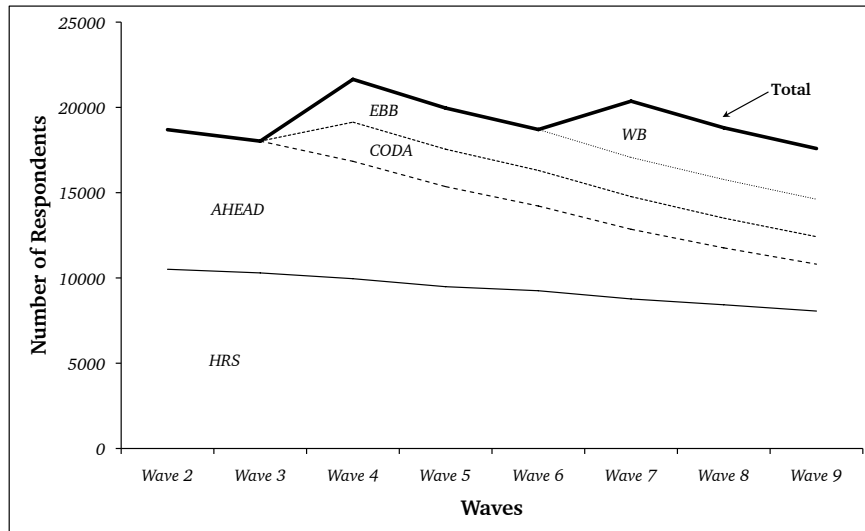
Given that the HRS survey study provides panel data that meets all of the above requirements, the present analysis keeps methodological changes to an absolute minimum and assesses the stability of HWW’s results when applying their model to new and more encompassing data.¹² Of particular interest is the question whether HWW’s somewhat surprising result of SES not having any direct causal impact on most health conditions is confirmed as test power increases.

4.1 The HRS panel data

Sample characteristics

¹²In fact, this study exactly replicates HWW’s model of health incidence with one notable exception. For simplicity, we skip their treatment of interview delay, which accounts for the fact that interview timing appears to depend on health status. While this potentially calls into question the comparability of responses from healthy and severely ill individuals, we find that results are virtually unaffected by this non-random distribution of time at risk.

Figure 1. Sample sizes, decomposed by entry cohorts



Our data – which is representative of the non-institutionalized US population over the age of 50 – comes from the *Health and Retirement Study* (HRS), a large-scale longitudinal survey project that studies the labor force participation and health transitions of individuals toward the end of their work lives and in the years that follow. While the data is collected by the University of Michigan Survey Research Center for the National Institute of Aging, we use the public-release file from the RAND Corporation that merged records from the nine panel waves available to date. The wave 1 interviews were conducted in 1992 and then repeated every two years, so that HRS incorporates data from 1992–2008. Due to significant changes to the survey design between waves 1 and 2, the first cross-section cannot be directly compared to subsequent observations and is therefore not used in our analysis. To ensure that HRS stays representative of the population as time goes by, the panel is periodically refreshed with new cohorts of respondents. Up to now, the sample consists of five different entry cohorts: the original 1992 “HRS” cohort (born 1931–1941), the 1993 “AHEAD” cohort (born 1923 or earlier), the “CODA” (born 1924–1930) and “WB” (born 1942–1947) cohorts entering in 1998, and the EBB cohort (born 1948–1953) added in 2004.

At baseline in wave 2 (covering interviews conducted between 1993 to 1994), the data set contains 18,694 individuals with usable records. The panel is subject to considerable attrition, which reduces sample sizes from wave to wave – a trend that is only temporarily disrupted when a refreshment cohort is added to the sample (see figure 1). The two sources for attrition are mortality (especially

for the elderly AHEAD cohort) and “sample fatigue”. Death-related attritors are kept in the working sample since mortality is one of the key outcomes of interest. With respect to all others attritors, we apply two alternative sampling schemes. The first exactly mirrors HWW’s benchmark in that it categorically excludes non-respondents from the working sample, irrespective of when their drop-off occurs or whether they rejoin the survey in later waves. As detailed below, the second sampling procedure assures that the information of these households is used for as long as they are part of the sample.

Much like in HWW’s original study, we exclude all individuals with missing information on critical variables. This includes item nonresponse for key demographic variables as well as cases where information on health conditions is generally unavailable. If respondents merely fail to answer isolated health queries, these gaps are filled by means of simulation-based imputation. Certain health questions on cognitive ability, severe falls, and hip fractures are not asked to participants below the age of 65, which is why these variables are excluded from all estimations that include younger sub-populations. While HWW went to great lengths to impute a large number of wealth and income observations with first-order Markov cross-wave hot deck imputation methods, we are in the more convenient position to rely on the imputations that are now readily available within the RAND/HRS data. We should note that, in spite of this data cleaning, the self-reported wealth and income measures are still suspect of considerable measurement error. The summary statistics for all variables used in our analysis is given in table [A-1](#).

Comparison with HWW’s data benchmark

HWW’s original data sample consists of the AHEAD cohort of US Americans aged 70 and older who are tracked through panel waves 2 to 4. Using the HRS data that is available to date, allows for deviations from this benchmark along several dimensions. Naturally, we can follow the same individuals for *more time periods* since the AHEAD cohort is now biennially observed between 1993/94 and 2008. Given the introduction of the four additional entry cohorts, the analysis can also be extended to *different individuals* with potentially diverging histories compared to those in the original study. In addition, it is now possible and certainly interesting to also widen the working sample by incorporating *younger individuals*, aged 50 and older. Finally, it should be noted that there is an additional, albeit minor, deviation from HWW’s data benchmark even for the same observations as in the original study. One reason for this is that the early AHEAD data has subsequently been subject to data updates and revisions within the HRS project. Similarly, there may be differences between the SES imputations carried out by HWW and those conducted by RAND/HRS.

4.2 Results

Following the strategy described in section 3, we fit model 9 as binomial probits except for BMI and ADL/IADL impairments, which are estimated with OLS and ordered probit, respectively. Appendix tables A-2 and A-3 contain the empirical significance values for the system of non-causality and invariance tests specified above. For a more concise overview of results, refer to tables 2 and 3. In a nutshell, the reanalysis with fresher and more encompassing data suggests that direct causal links from SES to health can be ruled out for much fewer health conditions than in the original study. This casts some doubt on the stability of HWW’s findings. In order to understand which of the data changes contribute to these deviating conclusions, we estimate the model multiple times, using several different data sets by augmenting them stepwise along the dimensions outlined above. In the first step, we rerun HWW’s benchmark study for the same cohort and time periods, yet reverting to the current version of HRS data instead.¹³ This will detect any impact arising from data revisions and differences in imputations. The second step consists of extending the analysis to the other three cohorts in the sample, hence testing whether HWW’s conclusions are also valid for different individuals. The third step addresses the question of how results are affected by increasing the number of time periods under consideration. Since in HWW’s model there is no self-evident way to aggregate the information from multiple time intervals, we compare two different sampling approaches: one that refills the sample after each wave with applicable observations and one that does not. In the fourth and final step, we evaluate the impact on estimation results when younger individuals are included in the analysis as well. This stepwise decomposition of all data and sampling changes appears to be more informative than applying the whole bundle of modifications at once.

Step one: Re-estimating HWW with revised data

In order to gauge the result’s sensitivity to data revisions and imputations, the model is re-estimated with fresh HRS data for the exact same cohort (AHEAD) and time periods (waves 2 to 4) as in the original study. The differences between the new results and HWW’s benchmark are quite modest, and outcomes of causality test are mostly unchanged. The most notable exception is diabetes

¹³To ensure that none of the observed changes is confoundingly rooted in the way certain variables are constructed and program codes are implemented, we also reran HWW’s study verbatim using their original data. While the goal to exactly reproduce HWW was ultimately achieved, it should be noted that results are identical to those published as log files within the appendix of the original 2003 paper, but not to those in the article itself. This difference is attributable to data revisions that HWW accounted for shortly after the paper was published, which means that the outcome from the appendix should be preferred as the ultimate benchmark. As is evident from comparing columns (1) and (2) of table A-2, said differences are not always trivial in size. Most strikingly, invariance tests tend to fail less frequently when applied to HWW’s post-publication data set. Yet, the impact on causality tests is negligible, thus not challenging the author’s main conclusions from the article.

Table 2. Tests for non-causality (Steps 2 to 4)

Health condition	Test results for Granger non-causality											
	HWW (70+)		Step 2 (70+)						Step 3 (70+)		Step 4 (50+)	
	W234		W234		W456		W789		W2-9		W2-9	
	F	M	F	M	F	M	F	M	F	M	F	M
Cancer									•	•		
Heart							•		•		••	••
Stroke					•						••	
Mortality									•	••	•••	•••
Lung		•		•								••
Diabetes		•									•••	
High bp.					•						•	
Arthritis							••		•••	••	•	
Incontinence					•		•	•	•••	••	•••	•••
Fall									•	•	n.a.	n.a.
Hip fract.							••				n.a.	n.a.
Proxy		•		•••		••			••	•••	•••	•••
Cognition	••	•	••		•	•	••	•	•••	•••	n.a.	n.a.
Psychiatric	••		•					•••		•	•••	•
Depression	•			••	••	••	•••	•	•••	•••	•••	•••
BMI									•		••	••
Smoke now					•						•••	•••
ADL	•		•						•••		•••	•••
IADL		•••		••				•••		•••	•••	•••
S.-r. health	••	•	•••	•	••	••	•••	•••	•••	•••	•••	•••

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Granger non-causality rejected at 5% level (•), rejected at 1% level (••), or rejected at 0.1% level (•••). Gray symbols indicate that the corresponding invariance test is rejected at the 5% level.

for which the non-causality test had to be previously rejected among male respondents. With revised data, however, a direct causal link from SES to diabetes seems unlikely to exist. For further details, compare columns (2) and (3) of table A-2.

Step two: Adding new cohorts

While the relative stability of results in face of data revisions is certainly encouraging, a much stricter test is posed by extending the analysis to all available cohorts. To achieve this, we run three separate estimations on the following samples: First, we revisit waves 2 to 4 but allow members of cohorts other than AHEAD to be part of the working sample. This barely changes the sample composition because the only other cohort that is part of the survey in this early stage is HRS, which hardly contains any individuals aged 70+. For the other two estimations, HWW's data benchmark is additionally changed inasmuch as later waves are used. The second estimation starts at wave 4, when the new cohorts CODA and EB are interviewed for the first time. Note that we do not restrict analysis to these two cohorts. Rather, all respondents who are at least 70 years old at wave 4 and who are not subject to subsequent sample attrition are followed until wave 7. This closely mirrors HWW's approach of analyzing a

three-period panel, hence still keeping the deviations from the benchmark to a minimum. The third estimation repeats the second for waves 7 to 9, coinciding with the entry of the most recent cohort, namely WB.

Not surprisingly, the first estimation (table A-2, column (4)) yields results that are almost identical to those of HWW's benchmark with revised data (table A-2, column (3)). The hypothesis tests associated with the other two estimations, however, prove to be rather different. As far as non-causality tests are concerned, the differences seem to be unsystematic. For some medical conditions, such as depression and ADL impairments for females and incontinence for males, causality from SES can no longer be ruled out as non-causality tests are now consistently rejected. For other health conditions, namely diabetes and lung disease for males as well as psychiatric disease for females, the opposite holds true, as non-causality tests can no longer be rejected. In addition, there are a number of diseases for which the benchmark causality test results are not confirmed for only one of the sub-samples. For further details compare columns (4), (5), and (6), respectively with column (3) of table A-2. Invariance tests, on the other hand, tend to be accepted more often than those under the benchmark scenario. At first glance, this may seem contradictory since causality tests have yielded fairly different results depending on which panel waves are under consideration. One should, however, not forget that the invariance tests merely check whether the model is time invariant within each of the three estimations but not among them. This is changed in the third step when the information from more than just three waves is incorporated.

Step three: Increasing the number of time periods

Step two has indicated that results depend on which panel waves are chosen to form the working sample. In order to reduce this arbitrary element and to maximize the use of available information in the data, it makes sense to increase the number of panel waves. Since there is no unequivocal way to implement this in practice, we propose two different sampling approaches. The first approach is a simple extrapolation of HWW's sampling method. The working sample consists of all individuals who participated in the survey in wave 2 and who were not subject to sample-fatigue-related attrition in later waves. This cohort is then followed for as many waves as possible. This sampling scheme has two major disadvantages. First, by restricting the sample to individuals who were part of the survey from the very start, we exclude refreshment cohorts CODA, EB, and WB, basically discarding useful information. The second drawback is of a more practical nature: death-related attritors cause the sample to dramatically thin out over time, so that sample sizes eventually become too small to conduct any meaningful analysis. Moreover, as time moves on, the sample arguably becomes less representative of the true population because the ongoing attrition

will select against the most frail. Nevertheless, and for the sake of maximum comparability with HWW, we estimate two versions of this first approach. One that follows individuals from wave 2 until wave 6 (covering cohorts HRS and AHEAD) and another that follows individuals from wave 4 until wave 9 (covering HRS, AHEAD, CODA, and EB). The number of waves is chosen so that sample sizes in the last respective wave are still reasonably large.

The alternative sampling scheme directly addresses the downsides of the approach above. Instead of limiting the sample to respondents who are part of the survey from the beginning, it is now refilled in each wave with all available respondents who meet the respective age criterion (i.e. 70+) and who answer all relevant questions for two consecutive waves. That way, all cohorts are used for analysis, sample sizes never diminish to levels too low for efficient estimation, and, consequently, all 8 waves can be used simultaneously. As a positive side effect, attrition bias is reduced as well, as the mortality-induced loss of observations is offset by filling up the sample with new respondents, once they become age-eligible. One might object that this approach reduces the power of panel analysis as it does not make much use of its potential time-series length. For the purpose of reproducing HWW, however, we deem it suitable since the original model does not use the theoretical length of the panel either, assuming that health and wealth trajectories are sufficiently described by single lags. Given that the models are estimated by simply stacking the data of all two-waves transitions above another, it is irrelevant how long an individual is part of the survey. The information conveyed in the responses of a person who only participates in, say, waves 4 and 5 is no less valuable than that of a respondent who participates from the very beginning to the end, and should therefore not be excluded.¹⁴

It is noteworthy that the invariance tests of both sampling schemes have slightly different interpretations. In both cases, they test whether parameter estimates stay constant over time, by comparing the log likelihood when all single wave transitions are pooled together with those when estimated separately. For a sample with refilling, an accepted invariance test indicates that the uncovered (non-)causal relationships hold for different populations at different times, underlining the generality of results. Invariance tests for samples without refilling, however, cannot answer the question whether causal links hold for different populations since only one cohort is being followed. They rather check whether these links remain constant as a steadily diminishing cohort becomes older over time, ultimately comparing the frail (who exit the sample early) with the medically more robust.

The main change in results from using a larger number of sample waves is that causality from SES to health can no longer be ruled out for a large array of

¹⁴Of course, the validity of this argument relies on HWW's conceptualization of health trajectories as a first-order Markov process. The future development of more realistic models of health dynamics will require a more sophisticated sampling procedure as well.

conditions, even for an elderly population, aged 70 and older. This observation holds, no matter which of the two above approaches is used, even though there are some differences. As columns (7), (8), and (9) of table A-2 reveal, there are seven health conditions for which the samples without refilling yield a rejection of non-causality tests, even though this was not the case for shorter samples. This number even increases to ten conditions, if the sample with refilling is used instead. The most unambiguous evidence exists for six conditions (mortality and falls for males, proxy and BMI for females, and cancer irrespective of gender) for which both approaches suggest that, contrary to earlier evidence, causality may well play a role. The reversed case of causality becoming less likely to exist as panel length grows, is as good as non-existent. The influence of analyzing more time periods at once on invariance tests is about the same for both approaches and not very strong. If anything, invariance failures tend to be somewhat likelier – a result that makes sense because it is more demanding for a model to be valid for eight waves than a mere three.

Step four: Adding younger individuals

So far, the consequence of applying HWW's model to data that includes more individuals and time periods, is that the number of medical conditions for which SES causality may play a role has considerably increased. However, for a population aged 70 and older there remains a large number of diseases for which causal links are not detected, despite the fact that high SES is associated with a lower prevalence of these conditions. While this cross-sectional correlation cannot be interpreted causally, it indicates that causal channels may have been at work earlier in life, before the individual even entered the sample. In light of that, it is interesting to also include younger individuals, to test if the data will pick up additional causal links that are already mute in an elderly population.

First, the sample is opened up to people who are at least 65 years old, so that it represents (with some exceptions) the whole Medicare-eligible subpopulation. This yields a net-increase of 3 to 6 health conditions (depending on whether samples with or without refilling are used) for which causality can no longer be rejected, affirming the speculation above. See columns (2), (6) and (10) of table A-3 for details. A similar effect can be observed when the sample is opened up even further to include individuals aged 50+, exploiting the entire age range available within HRS. This time, the net-increase amounts to another 3 to 9 conditions, rendering cases for which causal links can be ruled out the exception rather than the rule. Among the latter are illnesses such as strokes and high blood pressure for males, lung disease for females, and cancer for both men and women. For all other health conditions the existence of causal links cannot be refuted. For further details, consider columns (3), (7) and (11) of table A-3.

Table 3. Tests for non-causality (Pre- vs. post-retirement population)

Health condition	Test results for Granger non-causality			
	Step 4 (0-64)		Step 4 (65+)	
	W2-9		W2-9	
	F	M	F	M
Cancer				
Heart			•	
Stroke			•	
Mortality		••	•	••
Lung				•
Diabetes	•••		••	•
High bp.			•	
Arthritis			•••	•••
Incontinence	•••		•••	•••
Fall	n.a.	n.a.	•	•
Hip fract.	n.a.	n.a.		
Proxy		•••	•••	•••
Cognition	n.a.	n.a.	•••	•••
Psychiatric	•••			•
Depression	•••	•••	•••	•••
BMI	•		••	
Smoke now	•••	•••		
ADL	•••	•••	•••	•••
IADL	•••	•••	•	•••
S.-r. health	•••	•••	•••	•••

Notes: The same abbreviations as in table 2 apply.

It is also worthwhile to split up the sample into older (65+) and younger (50–64) individuals to study how the activation of causal channels differs between a mostly retired, Medicare-eligible population and people who are typically still on the labor market and not quasi-universally health insured. As table 3 shows, there is quite a number of medical conditions for which SES may be a causal driving force irrespective of age. These include depression for both genders, IADL impairments, incontinence, and diabetes for women as well as ADL impairments for men. For other conditions like arthritis, heart disease, strokes for females, or incontinence for males, SES is only a good predictor of new medical incidences at a higher age. On the other hand, smoking behavior as well as psychiatric problems for women are among the conditions for which a causal link may only be active at a pre-retirement age. Intriguingly, when young and old people are studied separately, results appear to be sensitive to whether samples with or without refilling are chosen (see table A-3). For older individuals, the sample with refilling suggests more cases of Granger causality than its counterparts without refreshment do. The exact opposite, however, is true for younger individuals as for these, rejected causality is a less frequent outcome in samples without refilling. The latter observation is likely an artifactual side effect of the way the sampling methods are defined: Sampling with

refilling effectively excludes people from the 50–64 sample once they become older than 65, whereas sampling without refilling follows all individuals until they die, even if they grow much older. As a consequence, unrefilled “young” samples may arguably pick up some of the causal effects that are exclusively active for the older subjects of the cohort.

Model invariance is not systematically influenced by adding younger individuals to the data set. The fact that the seeming structural breaks in the relation of SES and health as people grow older are not detected by HWW’s invariance tests, should, however, not be surprising. Recall that the test design does not pit the young against the old but the past against the future. The idea is to check parameter invariance as time progresses. Since the age structure within the sample varies only little from wave to wave (especially when it is regularly refreshed), the invariance test will not permit a direct comparison of, say, pre- and post-retirement populations. In light of this, the results in table A-3 merely suggest that the *time* stability of the model is rather insensitive to changes in the age composition of the sample.

Changes in results of the underlying prediction models

Given the strong dependence of non-causality test results on both the size and age coverage of the estimation sample, it seems natural to investigate how these changes are related to the size and the precision of coefficients of the underlying prediction models.¹⁵ As table 4 exemplifies, precision of SES coefficient estimates does generally increase with the size of the respective sample, even though this relation is not perfect. While standard errors remain fairly constant across estimations based on similarly sized 3-waves samples (step 2), they surprisingly spike upwards once all waves are pooled together in step 3. This observation may well be rooted in the aforementioned switch from a sampling procedure without (steps 1 and 2) to one with refilling (steps 3 and 4). Precision follows a more predictable pattern within step 4, as standard errors are invariably smaller, the larger the respective sample (note that $N_{50+} > N_{65+} > N_{70+} > N_{0-64}$).

Table 4. *Prediction model: Average standard errors of SES coefficients*

Sex	Standard errors of SES coefficients (average of all health cond. and SES regressors)							
	HWW	Step 2 (70+)			Step 3 (70+)	Step 4 (W2-9)		
	W234	W234	W456	W789	W2-9	65+	50+	0-64
Female	0.081	0.083	0.087	0.091	0.126	0.107	0.085	0.172
Male	0.108	0.105	0.111	0.110	0.167	0.135	0.098	0.185

Notes: Reported are average standard errors of SES coefficients obtained from estimating model 9. Each entry is an average of 160 single standard errors (20 health variables and 8 SES regressors). Individual standard errors follow the depicted pattern quite uniformly.

¹⁵HWW did not report the coefficients of the prediction models, but these estimates are also available in their online appendix.

Table 5. Prediction model: Significant SES coefficients

Health condition	SES regression coefficients that are significant at 5% level											
	HWW (70+)		Step 2 (70+)						Step 3 (70+)		Step 4 (50+)	
	W234		W234		W456		W789		W2-9		W2-9	
	F	M	F	M	F	M	F	M	F	M	F	M
Cancer		<i>E</i>		<i>E</i>			<i>H</i>		<i>W</i>			
Heart	I		I	<i>I</i>			E			NH		N
Stroke	<i>I</i>		<i>E</i>		<i>E</i>		W	<i>H</i>		I	W	
Mortality	<i>E</i>		<i>E</i>							W		W
Lung	W	W	W	W	H		<i>E</i>		WH		W	
Diabetes		E		E	W			W	E		WE	
High bp.	E			E	W			E				
Arthritis	E		E				I		W		<i>E</i>	
Incontinence	N				<i>E</i>	N	<i>I</i>	<i>E</i>	<i>I</i>		<i>W</i>	<i>H</i>
Fall							<i>E</i>	<i>E</i>			n.a.	n.a.
Hip fract.	W	I	W	I		W					n.a.	n.a.
Proxy	<i>I</i>	<i>E</i>	<i>I</i>	<i>E</i>		<i>E</i>	W		<i>I</i>		E	
Cognition	WIN	I	WIN	<i>I</i>	E		E		E		n.a.	n.a.
Psychiatric	I	W	I	W								
Depression		N		N	E	IH	EH		E	E	WE	
BMI					H		<i>IH</i>		E	E		H
Smoke now		W		W	W	N	<i>W</i>		W	E	W	I
ADL					<i>E</i>	<i>E</i>	<i>IE</i>				WI	W
IADL		WI		WIH	<i>IE</i>	N		E	H	W	N	W
S.-r. health	EH	WE	EH	WE	N	EH	INH	IEH	IEH	NH	WIENH	IEH

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Significant coefficients at 5% level for wealth (W), income (I), education (E), neighborhood (N), or housing quality (H). Bold (e.g. **W**) symbols indicate that signs are in line with theory. Italic (e.g. *W*) symbols denote the opposite case.

The same pattern emerges when comparing results by gender, as the number of women exceeds that of men in each of the subsamples.

Despite increased precision, table 5 suggests that the number of statistically significant SES coefficients does not seem to be systematically affected as samples become more encompassing. In HWW's benchmark sample, there is a total of 29 SES regression coefficients that are significant at the 5% level. This number stays rather constant across the performed steps. Yet, note that in the benchmark case there is a fair amount of cases for which coefficients have an unintuitive sign, suggesting that respondents with high SES are likelier to develop a new health condition. The share of these cases stays rather high for all 3-waves samples and drops significantly once all waves are used at once. This means that, while the overall number of significant SES coefficients does not increase, the direction of effects is now more in line with theory. This is an additional insight since non-causality tests as implemented here do merely check whether health innovations are conditionally independent of SES variables, whereas the quality of this dependence is not under consideration.

5 Conclusion and future research

All in all, re-estimating HWW’s model of health incidence with new HRS data, alters conclusions about SES causation quite significantly. While the impact of data revisions within HRS is encouragingly small, the addition of new cohorts shows that causal inference critically depends on which time periods are used for estimation. Using the information of many – ideally all – waves at once has the greatest effect on results, with many health conditions moving to the column of illnesses for which SES causality may well play a role. Adding younger individuals to the sample has a very similar effect, reducing the number of medical conditions for which the existence of causal links can be statistically rejected even further. As a consequence, the only health conditions for which SES causation can be ruled out when estimation is based on the most encompassing dataset with refilling, are cancer (irrespective of gender), lung disease for females and high blood pressure for males. For all other health incidences, SES is either G-causal or the failure of invariance tests does not permit reliable conclusions. This represents a stark contrast to HWW’s original findings, where the rejection of structural causality was the most frequent outcome.

Given that the greatest changes are triggered by the addition of panel waves (step 3), the main driving force behind this reversal in results is most likely an increase in test power as sample sizes soar. After all, in HWW’s stacking model, a longer panel is equivalent to a larger sample (with respect to N) since all waves are pooled together and treated as if they formed one cross section. This interpretation is corroborated by the fact that test results from long panels do not always reflect the average outcome of the respective three-wave panels they consist of. As the example of cancer in table A-2 illustrates, non-causality tests are often rejected in the long samples, even though they are consistently accepted in each of the short panels. Similar observations can be made when comparing test results by age group. In some cases, a non-causality test is only rejected for the largest, most-encompassing sample of all individuals aged 50 and older. However, in all smaller sub-samples (50–64, 65+, and 70+) the same null hypothesis cannot be rejected. As an example for the latter case, see heart disease for females in table A-3. All of this evidence permits the emergence of a clear picture: the larger the sample under consideration, the likelier the rejection of non-causality.

We also find that causal inference depends on the age structure of the underlying population, with certain conditions being Granger caused by SES at younger or older ages only. This yields at least indirect evidence that the activation of causal links may indeed change over the life cycle. However, we recommend to take these results with a grain of salt since their lack of robustness is far from comforting, as evidenced by the sensitivity to the choice of sampling

schemes. In addition, we should note that the data set for the 65+ population is about three times as large as that of the pre-retirement group. As a consequence, we face the risk of confounding the true effect of age structures with the impact of varying sample sizes identified above. This may well provide an alternative explanation for the failure to detect many cases of G-causality among the 50 to 64 year olds if estimation is based on a refilling sample – a result that is not confirmed if samples without refilling are used instead.

From a methodological point of view, the results of this study pose bad news for a model whose identification strategy relies on Granger causality. Recall that the reduced-form nature of G-causality cannot discriminate between structural causation and ecological association due to common unobserved effects if G-causality is detected. Ultimately, HWW's framework allows only the one-sided test for the absence of direct causal links, which is confirmed if G-causality is rejected as well. While HWW's original dataset provided us with a large number of such cases, the more-encompassing data samples analyzed here, do not do us this favor. As a result, we find ourselves in the unfortunate situation that little can be learned about the true links between SES and health, making it impossible to draw meaningful policy conclusions.¹⁶

In light of this, the need to improve the empirical model within future research so as to account for the confounding influence of hidden common factors, becomes even more pressing. In our view, there are two alternative ways to achieve this. The first mirrors the identification strategy of IV approaches: instead of using endogeneity-stricken SES histories as regressors, one could concentrate on the impact of clearly exogenous *changes* in these variables. If these SES innovations meet the standard IV assumptions, we would be able to formulate two-sided tests that permit the clean identification of causality in a structural sense. Among the natural experiments one could exploit, are the major negative shock to housing and financial wealth that many people experienced during the ongoing financial market crisis of 2008/09, the positive shock Medicare households received as a result of the introduction of the heavily subsidized Medicare Part D program in 2006, and the shocks some employed individuals received from changes in employer-provided health insurance.¹⁷ Particular atten-

¹⁶While invariance tests have arguably gained power by the inclusion of different time spans, cohorts, and age structures, we are still doubtful that their acceptance would attest the model the kind of stability necessary to make out-of-sample predictions of policy effects. The reason for this is that – with the notable exception of the introduction of Medicare Part D in 2006 – the observed variation in relevant policies remains rather low.

¹⁷When it comes to the recent financial crisis, we acknowledge that the equity shock might not be large enough to provide strong identification. Using HRS data, ? report that equity accounted only for about 15% of assets prior to the 2008/09 crisis. Whether this is sufficient exogenous variation would have to be scrutinized as part of future research. Alternatively, one could explore negative shocks to housing wealth which represent another aspect of the financial market crunch. Exogenous variation in these shocks is provided by regional differences in house prices and the severity of declines in real estate value during the crisis.

tion could be given to the differential exposure to wealth shocks in the presence of health care delivery systems that vary in the financial impact of copayments, premiums, and coverage, particularly for chronic conditions and preventative and palliative therapies. Provided that the causal link in questions even exists, wealth shocks will take some time to affect health outcomes. Therefore, we expect any effects of the 2008/2009 financial crisis or Medicare Part D to leave their marks only in future waves of the HRS dataset.

However, the use of such natural experiments is not immune to objection, which leads to a fundamental trade-off. On the one hand, we can try to infer causality by relying on wealth shocks like the ones just described, which has the advantage of not having to worry about endogeneity issues. Yet, as argued in section 3, there is a risk that these shocks may not be all that *relevant* for health, especially when occurring late in life. On the other hand, the information contained in past levels of SES – the regressor used in HWW’s G-causality framework – is certainly of great relevance, as it reflects the entire history between SES and health status. The disadvantage is that this pool of information may also include confounding elements, such as the impact of hidden common causes, calling into question the exogeneity of such explanatory variables.

The other alternative we deem feasible of discriminating among hypotheses A and C, seeks to solve this trade-off by exploiting the relevant information contained in SES histories, while eliminating the misleading influence of common effects. As extensively argued in the fixed- and random-effects literature, this may be achieved by interpreting the problem of common effects as an issue of unobserved individual heterogeneity, whose effect is controlled by fully exploiting the panel structure of HRS. This being said, the choice of a suitable estimator is not trivial because it needs to combine three important features that often tend to be mutually exclusive. First and foremost, the estimation strategy must allow heterogeneity to be correlated with SES, which makes FE estimators a logical candidate. However, FE estimation is generally ridden by matters of inconsistency, once confronted with the other two features, namely a dependent variable that is both binary (requiring a non-linear specification) and state-dependent (reflecting the dynamic nature of the model). A feasible way of tackling these three issues at once, promises to be a dynamic correlated RE Probit approach as implemented by ?. It solves the usual trade-offs between FE and RE setups, by allowing for correlated heterogeneity and the estimation of time-invariant regressors even when confronted with non-linear data structures and lagged dependent variables.¹⁸ We acknowledge that this alternative

¹⁸? adopt a similar strategy, by eliminating the effect of individual heterogeneity with GMM estimators in the spirit of ?. In analyzing the HRS population aged 51–61, they find that causal effects of wealth on health can be ruled out if unobserved heterogeneity and a more realistic lag structure are accounted for. However, given that their approach is incompatible with non-linear models, it is not directly applicable to our research question.

strategy of coping with common effects is not devoid of criticism either, which is why we consider it reasonable to independently explore both routes in what lies ahead. This is especially true inasmuch as both approaches are expected to uncover different causal channels: while the latter strategy of modelling individual heterogeneity may allow the detection of average causal effects as manifest in SES histories, the exploitation of natural experiments will predominantly shed light on the most immediate (mental) health consequences of wealth shocks.

A second opportunity for future research lies in improving the limited micro-foundation of causal pathways, which is inherent in the reduced-form nature of Granger causality. Even if we were able to univocally confirm the presence of causal effects from wealth to health, we still would not know the channels through which they operate. Yet, the latter information is absolutely critical from a policy perspective: interventions to increase the affordability of health insurance would be warranted if channel A1 were to be active, but would prove ineffective if the causal link were to work through, say, channel A3 instead. To address this issue, we intend to specify and test more differentiated hypotheses that may facilitate the discrimination among these channels. For instance, if channel A1 is truly relevant, we should observe a certain sensitivity of results to the availability and generosity of health care systems. Possible comparisons include the time before and after Medicare Part D, individuals with and without health insurance, or cross-country differences in health care regimes.¹⁹ Another way of gauging the importance of health care affordability is to compare individuals with and without health insurance. Of particular interest will be the pre-retirement population not yet eligible for Medicare, as their insurance status will be endogenous unless they are covered by employer-provided health care. Even if health insurance proves to be of little importance for the *onset* of a health condition, it may well be decisive in determining whether and how it is treated, given that the individual has already gotten sick. On this account we intend to follow the health trajectories as well medical care use of respondents that share the characteristic of having developed a certain medical condition.

Finally, the model would certainly benefit from addressing another of the methodological shortcomings identified in section 3: the treatment of health dynamics. In our view, there are several ways to accommodate the long memory effects that prove to be so critical for a realistic description of health trajectories. The simplest fix consists of adding higher-order lags of health condition prevalences to the list of explanatory variables. A better, albeit more demanding, alternative is a hidden Markov structure in which health is controlled by

¹⁹In fact, we find that causal effects from SES to health status are less pronounced in the Netherlands than in the USA. Given that the Dutch health care system is basically universal, they see this result as an indication of the general importance of differential access to health care: SES gradients in health are strongest in institutional environments in which affordability should a priori matter most.

a latent random process that drives the onset of health conditions, self-rated health and mortality. According to ?, such models are parsimonious and capture the observed dynamics better than commonly applied random-effects or conditional Markov chain models.

Table A-1. Variables used for analysis: Summary statistics

Variable	Wave 2 (N=18,694)		Wave 3 (N=18,022)		Wave 4 (N=21,645)		Wave 5 (N=19,961)		Wave 6 (N=18,703)		Wave 7 (N=20,365)		Wave 8 (N=18,806)		Wave 9 (N=17,588)	
	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.
<i>Health prevalence</i>																
Cancer (ever)	0.097	0.296	0.127	0.333	0.127	0.333	0.140	0.347	0.155	0.362	0.152	0.359	0.168	0.373	0.179	0.384
Heart disease (ever)	0.225	0.418	0.265	0.442	0.256	0.436	0.270	0.444	0.287	0.452	0.275	0.447	0.296	0.456	0.306	0.461
Stroke (ever)	0.066	0.248	0.090	0.286	0.090	0.287	0.098	0.297	0.105	0.306	0.099	0.298	0.107	0.309	0.113	0.316
Lung disease (ever)	0.107	0.309	0.115	0.319	0.102	0.302	0.103	0.304	0.106	0.308	0.105	0.307	0.112	0.316	0.120	0.325
Diabetes (ever)	0.133	0.340	0.150	0.357	0.140	0.357	0.161	0.368	0.183	0.387	0.187	0.390	0.211	0.408	0.229	0.420
High blood pressure (ever)	0.464	0.499	0.500	0.500	0.488	0.500	0.517	0.500	0.556	0.497	0.553	0.497	0.597	0.491	0.629	0.483
Arthritis (last 2 years)	0.190	0.392	0.227	0.419	0.220	0.414	0.223	0.416	0.237	0.425	0.228	0.420	0.231	0.421	0.240	0.427
Incontinence (ever)	0.146	0.353	0.214	0.410	0.237	0.425	0.290	0.454	0.331	0.471	0.324	0.468	0.364	0.481	0.398	0.490
Fall (ever)	0.075	0.263	0.148	0.355	0.169	0.374	0.203	0.402	0.225	0.418	0.247	0.431	0.262	0.440	0.288	0.453
Hip fracture (ever)	0.024	0.152	0.025	0.157	0.027	0.162	0.029	0.167	0.031	0.173	0.035	0.183	0.038	0.192	0.040	0.195
Proxy interview (now)	0.077	0.267	0.138	0.345	0.145	0.352	0.164	0.370	0.182	0.386	0.143	0.350	0.131	0.337	0.134	0.341
Cognitive impairment (ever)	0.221	0.415	0.151	0.358	0.208	0.406	0.257	0.437	0.254	0.435	0.198	0.398	0.234	0.424	0.230	0.421
Psychiatric disease (ever)	0.128	0.334	0.158	0.365	0.151	0.358	0.159	0.365	0.170	0.375	0.177	0.381	0.191	0.393	0.199	0.400
Depression (last 2 years)	0.130	0.337	0.096	0.295	0.110	0.313	0.111	0.314	0.112	0.315	0.111	0.314	0.114	0.318	0.104	0.305
BMI (now)	26.5	5.0	26.5	5.1	26.8	5.2	27.0	5.4	27.2	5.4	27.4	5.7	27.8	5.8	28.0	5.9
Smoker (now)	0.182	0.386	0.169	0.375	0.164	0.371	0.149	0.356	0.137	0.343	0.148	0.355	0.136	0.343	0.133	0.339
# of ADL impairments (now)	0.239	0.751	0.378	0.990	0.376	0.998	0.384	1.003	0.393	1.020	0.364	0.984	0.399	1.022	0.403	1.035
# of IADL impairments (now)	0.288	0.847	0.344	0.967	0.336	0.966	0.345	0.990	0.370	1.027	0.340	0.964	0.369	1.014	0.377	1.027
Poor/fair self-rated h. (now)	0.291	0.454	0.282	0.450	0.319	0.466	0.286	0.452	0.294	0.456	0.300	0.458	0.301	0.459	0.307	0.461
<i>Health incidence</i>																
Cancer (first/new)			0.040	0.197	0.046	0.209	0.039	0.194	0.044	0.206	0.041	0.198	0.042	0.201	0.046	0.209
Heart disease (first/new)			0.092	0.289	0.125	0.331	0.099	0.299	0.110	0.313	0.117	0.322	0.107	0.309	0.110	0.313
Stroke (first/new)			0.038	0.191	0.047	0.213	0.038	0.190	0.042	0.201	0.039	0.194	0.036	0.186	0.039	0.193
Died since last wave			0.054	0.225	0.055	0.229	0.065	0.246	0.076	0.264	0.058	0.234	0.066	0.249	0.071	0.257
Lung disease (first)			0.017	0.127	0.020	0.139	0.017	0.130	0.020	0.141	0.023	0.149	0.018	0.133	0.021	0.142
Diabetes (first)			0.023	0.151	0.023	0.150	0.023	0.150	0.032	0.176	0.028	0.166	0.032	0.176	0.030	0.171
High blood pressure (first)			0.043	0.204	0.045	0.207	0.050	0.218	0.059	0.235	0.054	0.227	0.055	0.227	0.052	0.221
Arthritis (first/new)			0.105	0.307	0.099	0.298	0.098	0.298	0.110	0.313	0.109	0.312	0.099	0.298	0.106	0.308
Incontinence (first/new)			0.159	0.365	0.166	0.372	0.183	0.387	0.201	0.401	0.201	0.401	0.220	0.414	0.252	0.440
Fall (first/new)			0.110	0.313	0.109	0.312	0.101	0.302	0.104	0.306	0.109	0.312	0.112	0.316	0.126	0.332
Hip fracture (first/new)			0.008	0.089	0.013	0.112	0.009	0.094	0.013	0.111	0.014	0.119	0.015	0.121	0.016	0.124

Proxy interview (now)			0.138	0.345	0.145	0.352	0.164	0.370	0.182	0.386	0.143	0.350	0.131	0.337	0.134	0.341
Cognitive impairment (first)			0.096	0.295	0.074	0.261	0.058	0.233	0.062	0.240	0.047	0.212	0.047	0.212	0.482	0.214
Psychiatric disease (first)			0.032	0.175	0.027	0.162	0.024	0.152	0.027	0.162	0.025	0.157	0.028	0.165	0.022	0.147
Depression (first/new)			0.050	0.218	0.072	0.258	0.061	0.240	0.062	0.242	0.059	0.236	0.060	0.238	0.053	0.223
BMI (now)			26.5	5.1	26.8	5.2	27.0	5.4	27.2	5.4	27.4	5.7	27.8	5.8	28.0	5.9
Smoker (now)			0.169	0.375	0.164	0.371	0.149	0.356	0.137	0.343	0.148	0.355	0.136	0.343	0.133	0.339
# of ADL impairments (now)			0.378	0.990	0.376	0.998	0.384	1.003	0.393	1.020	0.364	0.984	0.399	1.022	0.403	1.035
# of IADL impairments (now)			0.344	0.967	0.336	0.966	0.345	0.990	0.370	1.027	0.340	0.964	0.369	1.014	0.377	1.027
Poor/fair self-rated h. (now)			0.282	0.450	0.319	0.466	0.286	0.452	0.294	0.456	0.300	0.458	0.301	0.459	0.307	0.461
<i>SES variables</i>																
Wealth in 1997 USD (000)	225.9	469.6	273.8	704.3	288.2	1,057	315.7	853.1	310.0	789.0	348.6	1,226	418.8	1,877	360.6	1,034
1st quartile wealth indicator	0.264	0.441	0.246	0.431	0.253	0.435	0.239	0.427	0.229	0.421	0.246	0.430	0.239	0.426	0.240	0.427
4th quartile wealth indicator	0.190	0.393	0.215	0.411	0.236	0.425	0.261	0.439	0.275	0.447	0.283	0.451	0.311	0.463	0.304	0.450
Income in 1997 USD (000)	44.1	79.5	45.2	65.9	47.2	90.2	47.7	93.1	45.0	80.8	50.1	85.3	50.6	255.4	49.0	359.1
1st quartile income indicator	0.289	0.453	0.268	0.443	0.251	0.434	0.253	0.435	0.250	0.433	0.240	0.427	0.242	0.428	0.251	0.434
4th quartile income indicator	0.213	0.410	0.230	0.421	0.247	0.432	0.249	0.432	0.236	0.425	0.263	0.441	0.249	0.432	0.253	0.435
Poor/fair housing condition	0.108	0.310	0.119	0.324	0.113	0.317	0.105	0.307	0.095	0.294	0.124	0.330	0.114	0.318	0.109	0.311
Poor/fair neighborhood safety	0.146	0.353	0.122	0.327	0.101	0.301	0.087	0.282	0.078	0.268	0.097	0.296	0.101	0.301	0.096	0.295
Education (in years)	11.5	3.5	11.6	3.5	11.8	3.4	11.9	3.4	12.0	3.4	12.3	3.3	12.3	3.3	12.4	3.3
High school (educ. > 10 y.)	0.701	0.458	0.706	0.456	0.737	0.440	0.746	0.435	0.758	0.429	0.787	0.410	0.795	0.404	0.801	0.400
College (educ. > 14 y.)	0.181	0.385	0.184	0.387	0.204	0.403	0.211	0.408	0.218	0.413	0.242	0.428	0.244	0.429	0.251	0.434
<i>Demographic variables</i>																
Widowed	0.200	0.400	0.209	0.407	0.203	0.402	0.216	0.412	0.230	0.421	0.201	0.401	0.213	0.410	0.221	0.415
Divorced/separated	0.087	0.282	0.087	0.281	0.095	0.293	0.095	0.294	0.097	0.296	0.109	0.311	0.109	0.312	0.114	0.318
Never married	0.031	0.174	0.029	0.168	0.033	0.177	0.031	0.174	0.030	0.170	0.037	0.188	0.034	0.182	0.034	0.180
Current age	65.9	11.0	67.2	10.7	66.5	10.9	67.7	10.6	69.0	10.3	67.0	11.5	68.4	11.1	69.6	10.8
Mother's current/death age	73.7	15.3	74.4	15.3	75.0	15.0	75.4	15.1	75.8	15.2	75.6	14.7	76.1	14.8	76.5	14.8
Father's current/death age	70.8	14.4	71.2	14.6	71.4	14.4	71.7	14.5	71.9	14.5	71.9	14.2	71.1	14.3	71.4	14.4
Ever smoked	0.588	0.492	0.587	0.492	0.590	0.492	0.586	0.493	0.583	0.493	0.579	0.494	0.579	0.494	0.577	0.494

Table A-2. Tests for invariance and non-causality (Ages 70+)

Health condition	Test	Sex	Significance levels of tests								
			Benchmark: HWW (70+)			3-waves samples (70+)			Longer samples (70+)		
			Article	Appendix	New data	No refilling			No refilling		Refilling
			W234	W234	W234	W234	W456	W789	W2-6	W4-9	W2-9
			(AHEAD)	(AHEAD)	(AHEAD)	(All Coh.)	(All Coh.)	(All Coh.)	(All Coh.)	(All Coh.)	(All Coh.)
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cancer	I	F	0.056	0.528	0.933	0.930	0.870	0.605	0.683	0.777	0.591
	I noS	F	0.023	0.309	0.895	0.896	0.845	0.543	0.628	0.788	0.489
	S I	F	0.311	0.613	0.271	0.264	0.483	0.280	0.350	0.020	0.048
	I&S	F	0.057	0.600	0.855	0.846	0.854	0.515	0.653	0.546	0.441
	I	M	0.000	0.016	0.119	0.172	0.355	0.179	0.349	0.585	0.156
	I noS	M	0.000	0.007	0.095	0.134	0.158	0.450	0.187	0.419	0.263
	S I	M	0.225	0.198	0.203	0.178	0.806	0.302	0.047	0.294	0.030
	I&S	M	0.000	0.012	0.088	0.117	0.516	0.163	0.194	0.545	0.081
Heart disease	I	F	0.000	0.085	0.000	0.000	0.849	0.392	0.017	0.874	0.051
	I noS	F	0.000	0.052	0.000	0.000	0.903	0.440	0.010	0.946	0.084
	S I	F	0.812	0.398	0.068	0.059	0.701	0.010	0.301	0.406	0.020
	I&S	F	0.000	0.099	0.000	0.000	0.891	0.080	0.016	0.863	0.021
	I	M	0.017	0.690	0.505	0.603	0.248	0.635	0.484	0.262	0.185
	I noS	M	0.020	0.619	0.458	0.566	0.197	0.856	0.384	0.702	0.343
	S I	M	0.290	0.243	0.162	0.224	0.468	0.533	0.216	0.077	0.098
	I&S	M	0.018	0.571	0.349	0.478	0.279	0.662	0.413	0.168	0.130
Stroke	I	F	0.068	0.384	0.016	0.015	0.814	0.690	0.023	0.710	0.206
	I noS	F	0.034	0.266	0.030	0.028	0.817	0.578	0.035	0.791	0.313
	S I	F	0.056	0.657	0.193	0.147	0.044	0.197	0.062	0.311	0.065
	I&S	F	0.023	0.481	0.012	0.009	0.461	0.540	0.010	0.675	0.135
	I	M	0.086	0.204	0.030	0.017	0.403	0.446	0.036	0.745	0.015
	I noS	M	0.042	0.135	0.006	0.003	0.276	0.334	0.004	0.437	0.014
	S I	M	0.290	0.059	0.213	0.195	0.750	0.664	0.215	0.529	0.682
	I&S	M	0.080	0.080	0.024	0.013	0.535	0.543	0.029	0.757	0.020
Mortality	I	F	0.010	0.221	0.092	0.085	0.169	0.517	0.165	0.491	0.011
	I noS	F	0.006	0.315	0.161	0.162	0.220	0.403	0.321	0.355	0.010
	S I	F	0.812	0.652	0.331	0.309	0.453	0.108	0.248	0.092	0.029
	I&S	F	0.030	0.308	0.093	0.083	0.195	0.312	0.140	0.368	0.005

	I	M	0.021	0.378	0.306	0.262	0.846	0.370	0.491	0.156	0.077
	I noS	M	0.032	0.492	0.236	0.227	0.594	0.395	0.318	0.064	0.072
	S I	M	0.228	0.364	0.274	0.247	0.371	0.154	0.331	0.002	0.009
	I&S	M	0.018	0.360	0.259	0.212	0.796	0.244	0.462	0.033	0.027
Lung disease	I	F	0.493	0.552	0.386	0.403	0.154	0.526	0.421	0.253	0.496
	I noS	F	0.479	0.483	0.486	0.487	0.154	0.324	0.435	0.273	0.732
	S I	F	0.381	0.343	0.218	0.190	0.669	0.057	0.185	0.219	0.329
	I&S	F	0.470	0.504	0.295	0.291	0.235	0.250	0.343	0.213	0.474
	I	M	0.603	0.689	0.210	0.273	0.059	0.493	0.102	0.069	0.241
	I noS	M	0.620	0.502	0.064	0.094	0.119	0.428	0.082	0.074	0.192
	S I	M	0.013	0.010	0.023	0.026	0.084	0.617	0.161	0.584	0.255
	I&S	M	0.174	0.203	0.052	0.076	0.026	0.569	0.073	0.085	0.215
Diabetes	I	F	0.110	0.189	0.344	0.379	0.572	0.667	0.158	0.282	0.384
	I noS	F	0.177	0.281	0.407	0.403	0.477	0.655	0.098	0.143	0.593
	S I	F	0.246	0.110	0.657	0.601	0.086	0.672	0.483	0.134	0.071
	I&S	F	0.091	0.101	0.439	0.454	0.329	0.740	0.174	0.211	0.280
	I	M	0.243	0.234	0.377	0.269	0.603	0.393	0.029	0.103	0.063
	I noS	M	0.199	0.128	0.284	0.179	0.585	0.403	0.021	0.061	0.068
	S I	M	0.043	0.025	0.188	0.149	0.609	0.304	0.794	0.419	0.115
	I&S	M	0.085	0.062	0.271	0.171	0.663	0.345	0.049	0.107	0.044
High blood pressure	I	F	0.004	0.007	0.008	0.010	0.154	0.217	0.006	0.683	0.099
	I noS	F	0.009	0.024	0.043	0.047	0.051	0.138	0.018	0.416	0.074
	S I	F	0.777	0.534	0.488	0.516	0.018	0.465	0.835	0.084	0.228
	I&S	F	0.012	0.013	0.014	0.018	0.032	0.246	0.012	0.550	0.084
	I	M	0.220	0.393	0.074	0.083	0.053	0.759	0.049	0.111	0.186
	I noS	M	0.121	0.172	0.014	0.017	0.029	0.579	0.029	0.041	0.052
	S I	M	0.668	0.990	0.697	0.633	0.474	0.283	0.327	0.035	0.091
	I&S	M	0.310	0.666	0.132	0.133	0.072	0.667	0.047	0.052	0.130
Arthritis	I	F	0.041	0.046	0.093	0.087	0.270	0.928	0.363	0.399	0.053
	I noS	F	0.017	0.032	0.104	0.109	0.156	0.663	0.234	0.352	0.015
	S I	F	0.042	0.085	0.034	0.055	0.622	0.003	0.011	0.035	0.000
	I&S	F	0.012	0.020	0.025	0.031	0.349	0.342	0.145	0.240	0.005
	I	M	0.187	0.071	0.081	0.102	0.259	0.960	0.125	0.581	0.884
	I noS	M	0.276	0.167	0.110	0.151	0.246	0.963	0.112	0.325	0.690
	S I	M	0.145	0.395	0.382	0.313	0.472	0.409	0.404	0.276	0.009
	I&S	M	0.116	0.082	0.090	0.098	0.289	0.935	0.128	0.538	0.718

Incontinence	I	F	0.357	0.781	0.285	0.271	0.494	0.798	0.195	0.540	0.185
	I noS	F	0.329	0.597	0.235	0.238	0.646	0.793	0.199	0.727	0.402
	S I	F	0.080	0.163	0.082	0.077	0.010	0.015	0.004	0.000	0.000
	I&S	F	0.183	0.615	0.142	0.130	0.126	0.351	0.048	0.134	0.000
	I	M	0.161	0.351	0.681	0.614	0.257	0.236	0.553	0.022	0.393
	I noS	M	0.486	0.682	0.924	0.903	0.396	0.337	0.730	0.312	0.616
	S I	M	0.237	0.463	0.296	0.374	0.316	0.010	0.175	0.048	0.002
	I&S	M	0.129	0.373	0.602	0.578	0.234	0.043	0.464	0.010	0.169
Fall	I	F	0.864	0.904	0.616	0.626	0.100	0.955	0.089	0.444	0.359
	I noS	F	0.763	0.790	0.372	0.384	0.193	0.997	0.037	0.525	0.478
	S I	F	0.515	0.600	0.494	0.482	0.507	0.064	0.296	0.261	0.015
	I&S	F	0.861	0.913	0.630	0.634	0.131	0.764	0.081	0.402	0.205
	I	M	0.402	0.263	0.126	0.248	0.044	0.549	0.013	0.275	0.301
	I noS	M	0.383	0.207	0.092	0.159	0.028	0.481	0.005	0.111	0.073
	S I	M	0.507	0.592	0.347	0.474	0.338	0.103	0.009	0.079	0.015
	I&S	M	0.438	0.329	0.126	0.278	0.047	0.342	0.003	0.185	0.164
Hip fracture	I	F	0.604	0.492	0.260	0.256	0.711	0.790	0.083	0.569	0.143
	I noS	F	0.470	0.280	0.283	0.284	0.646	0.907	0.086	0.302	0.043
	S I	F	0.275	0.159	0.275	0.259	0.597	0.983	0.195	0.821	0.689
	I&S	F	0.520	0.348	0.222	0.213	0.750	0.925	0.064	0.643	0.170
	I	M	0.056	0.126	0.010	0.017	0.359	0.104	0.000	0.097	0.006
	I noS	M	0.051	0.053	0.105	0.111	0.203	0.094	0.036	0.155	0.006
	S I	M	0.305	0.430	0.069	0.078	0.003	0.686	0.011	0.048	0.203
	I&S	M	0.055	0.143	0.004	0.007	0.048	0.167	0.000	0.051	0.005
Proxy interview	I	F	0.345	0.301	0.597	0.582	0.283	0.764	0.953	0.482	0.444
	I noS	F	0.442	0.465	0.922	0.922	0.506	0.663	0.998	0.782	0.652
	S I	F	0.250	0.254	0.130	0.144	0.236	0.134	0.009	0.182	0.001
	I&S	F	0.283	0.248	0.413	0.412	0.226	0.577	0.784	0.414	0.185
	I	M	0.424	0.591	0.013	0.018	0.040	0.642	0.000	0.001	0.001
	I noS	M	0.326	0.378	0.009	0.013	0.017	0.884	0.000	0.001	0.010
	S I	M	0.019	0.032	0.000	0.000	0.002	0.925	0.000	0.004	0.000
	I&S	M	0.131	0.261	0.000	0.000	0.001	0.803	0.000	0.000	0.000
Cognitive impairment	I	F	0.020	0.005	0.278	0.287	0.417	0.057	0.445	0.053	0.090
	I noS	F	0.007	0.008	0.436	0.435	0.239	0.108	0.430	0.013	0.127
	S I	F	0.001	0.002	0.002	0.002	0.034	0.007	0.052	0.000	0.000
	I&S	F	0.000	0.000	0.027	0.026	0.160	0.006	0.285	0.004	0.001

	I	M	0.429	0.288	0.129	0.167	0.357	0.772	0.219	0.034	0.157
	I noS	M	0.245	0.143	0.068	0.113	0.408	0.705	0.166	0.052	0.228
	S I	M	0.022	0.026	0.049	0.060	0.031	0.010	0.036	0.001	0.000
	I&S	M	0.140	0.089	0.046	0.068	0.126	0.291	0.109	0.005	0.003
Psychiatric disease	I	F	0.075	0.127	0.164	0.154	0.711	0.075	0.074	0.945	0.175
	I noS	F	0.031	0.052	0.147	0.150	0.491	0.296	0.063	0.950	0.269
	S I	F	0.012	0.004	0.012	0.011	0.270	0.522	0.041	0.737	0.299
	I&S	F	0.012	0.012	0.031	0.028	0.622	0.102	0.033	0.958	0.162
	I	M	0.194	0.295	0.036	0.037	0.009	0.099	0.066	0.006	0.088
	I noS	M	0.546	0.365	0.106	0.115	0.028	0.174	0.212	0.021	0.143
	S I	M	0.110	0.065	0.110	0.080	0.674	0.004	0.428	0.076	0.035
	I&S	M	0.108	0.137	0.019	0.016	0.019	0.009	0.071	0.003	0.048
Depression	I	F	0.299	0.211	0.196	0.197	0.236	0.125	0.199	0.732	0.068
	I noS	F	0.347	0.299	0.503	0.505	0.182	0.133	0.177	0.748	0.095
	S I	F	0.078	0.011	0.086	0.107	0.001	0.000	0.000	0.000	0.000
	I&S	F	0.151	0.042	0.098	0.109	0.019	0.002	0.023	0.062	0.000
	I	M	0.767	0.944	0.578	0.694	0.185	0.252	0.007	0.452	0.093
	I noS	M	0.856	0.934	0.478	0.581	0.209	0.125	0.007	0.414	0.041
	S I	M	0.302	0.065	0.003	0.002	0.005	0.016	0.000	0.007	0.000
	I&S	M	0.695	0.753	0.112	0.148	0.023	0.062	0.000	0.222	0.002
BMI	I	F	0.419	0.260	0.192	0.192	0.113	0.008	0.039	0.038	0.001
	I noS	F	0.330	0.251	0.096	0.104	0.127	0.025	0.018	0.027	0.000
	S I	F	0.738	0.673	0.932	0.953	0.352	0.083	0.503	0.001	0.011
	I&S	F	0.531	0.345	0.356	0.368	0.115	0.003	0.046	0.006	0.000
	I	M	0.010	0.009	0.011	0.015	0.000	0.353	0.000	0.000	0.000
	I noS	M	0.002	0.002	0.014	0.017	0.001	0.618	0.000	0.000	0.000
	S I	M	0.249	0.664	0.664	0.623	0.642	0.248	0.931	0.316	0.487
	I&S	M	0.009	0.019	0.022	0.028	0.000	0.290	0.000	0.000	0.000
Smoke now	I	F	0.509	0.420	0.151	0.158	0.267	0.760	0.085	0.421	0.031
	I noS	F	0.242	0.214	0.063	0.063	0.382	0.763	0.062	0.655	0.046
	S I	F	0.838	0.625	0.637	0.635	0.034	0.082	0.140	0.309	0.172
	I&S	F	0.650	0.492	0.210	0.219	0.097	0.525	0.060	0.396	0.024
	I	M	0.366	0.710	0.598	0.579	0.001	0.128	0.000	0.577	0.221
	I noS	M	0.146	0.504	0.390	0.420	0.068	0.196	0.679	0.016	0.374
	S I	M	0.064	0.129	0.217	0.424	0.539	0.721	0.459	0.400	0.305
	I&S	M	0.182	0.528	0.488	0.570	0.001	0.200	0.000	0.569	0.208

ADL impairment	I	F	0.010	0.004	0.020	0.019	0.518	0.133	0.046	0.193	0.000
	I noS	F	0.018	0.022	0.047	0.048	0.343	0.156	0.031	0.181	0.000
	S I	F	0.818	0.831	0.859	0.850	0.011	0.049	0.238	0.000	0.000
	I&S	F	0.027	0.011	0.049	0.046	0.170	0.052	0.039	0.042	0.000
	I	M	0.004	0.067	0.016	0.020	0.339	0.229	0.026	0.397	0.018
	I noS	M	0.044	0.114	0.018	0.026	0.692	0.640	0.305	0.468	0.264
	S I	M	0.370	0.352	0.093	0.131	0.232	0.068	0.453	0.639	0.081
	I&S	M	0.005	0.069	0.008	0.012	0.275	0.111	0.029	0.433	0.011
IADL impairment	I	F	0.673	0.724	0.615	0.598	0.207	0.276	0.467	0.108	0.442
	I noS	F	0.514	0.525	0.555	0.544	0.150	0.296	0.312	0.106	0.337
	S I	F	0.368	0.230	0.303	0.281	0.021	0.668	0.114	0.110	0.053
	I&S	F	0.636	0.625	0.555	0.529	0.061	0.353	0.364	0.075	0.335
	I	M	0.016	0.024	0.038	0.022	0.658	0.202	0.009	0.137	0.053
	I noS	M	0.003	0.011	0.024	0.014	0.633	0.069	0.003	0.084	0.010
	S I	M	0.006	0.009	0.006	0.004	0.127	0.009	0.023	0.001	0.000
	I&S	M	0.002	0.003	0.004	0.002	0.486	0.041	0.003	0.031	0.002
Self-rated health	I	F	0.151	0.376	0.006	0.008	0.021	0.493	0.004	0.000	0.000
	I noS	F	0.111	0.341	0.018	0.019	0.078	0.358	0.024	0.000	0.000
	S I	F	0.001	0.001	0.000	0.000	0.004	0.000	0.000	0.000	0.000
	I&S	F	0.009	0.044	0.000	0.000	0.004	0.001	0.000	0.000	0.000
	I	M	0.581	0.634	0.733	0.650	0.033	0.062	0.052	0.039	0.004
	I noS	M	0.570	0.607	0.610	0.574	0.008	0.033	0.019	0.009	0.006
	S I	M	0.034	0.020	0.004	0.003	0.002	0.000	0.000	0.000	0.000
	I&S	M	0.282	0.278	0.226	0.169	0.002	0.000	0.001	0.000	0.000

Notes: Results are for white females (F) and males (M). Abbreviations of tests are as follows: I = Unconditional invariance; I|noS = Invariance, conditional on non-causality; S|I = Non-causality, conditional on invariance; I&S = Joint invariance and non-causality.

Table A-3. Tests for invariance and non-causality (All ages)

Health condition	Test	Sex	Significance levels of tests											
			No refilling								Refilling			
			W2-6				W4-9				W2-9			
			70+	65+	50+	50-64	70+	65+	50+	50-64	70+	65+	50+	50-64
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Cancer	I	F	0.683	0.666	0.889	0.126	0.777	0.506	0.265	0.388	0.591	0.250	0.497	0.410
	I noS	F	0.628	0.454	0.720	0.125	0.788	0.513	0.238	0.306	0.489	0.078	0.281	0.318
	S I	F	0.350	0.367	0.659	0.552	0.020	0.007	0.064	0.349	0.048	0.151	0.077	0.301
	I&S	F	0.653	0.642	0.909	0.151	0.546	0.229	0.159	0.370	0.441	0.198	0.375	0.383
	I	M	0.349	0.269	0.160	0.021	0.585	0.623	0.136	0.108	0.156	0.024	0.116	0.068
	I noS	M	0.187	0.096	0.245	0.056	0.419	0.647	0.119	0.142	0.263	0.034	0.112	0.055
	S I	M	0.047	0.082	0.179	0.993	0.294	0.438	0.315	0.689	0.030	0.109	0.110	0.917
	I&S	M	0.194	0.166	0.120	0.058	0.545	0.620	0.127	0.143	0.081	0.015	0.080	0.106
Heart disease	I	F	0.017	0.012	0.043	0.113	0.874	0.867	0.853	0.397	0.051	0.057	0.211	0.546
	I noS	F	0.010	0.007	0.028	0.144	0.946	0.864	0.912	0.749	0.084	0.039	0.112	0.520
	S I	F	0.301	0.192	0.035	0.679	0.406	0.053	0.041	0.709	0.020	0.020	0.001	0.115
	I&S	F	0.016	0.009	0.016	0.153	0.863	0.733	0.689	0.460	0.021	0.023	0.053	0.448
	I	M	0.484	0.380	0.405	0.601	0.262	0.336	0.558	0.523	0.185	0.109	0.480	0.117
	I noS	M	0.384	0.241	0.149	0.376	0.702	0.569	0.485	0.474	0.343	0.200	0.343	0.056
	S I	M	0.216	0.316	0.071	0.045	0.077	0.092	0.025	0.043	0.098	0.084	0.007	0.119
	I&S	M	0.413	0.352	0.255	0.384	0.168	0.234	0.333	0.336	0.130	0.070	0.245	0.083
Stroke	I	F	0.023	0.008	0.018	0.073	0.710	0.690	0.319	0.199	0.206	0.152	0.201	0.008
	I noS	F	0.035	0.043	0.052	0.046	0.791	0.730	0.377	0.138	0.313	0.248	0.286	0.037
	S I	F	0.062	0.167	0.007	0.038	0.311	0.195	0.051	0.252	0.065	0.027	0.007	0.332
	I&S	F	0.010	0.005	0.003	0.029	0.675	0.616	0.189	0.172	0.135	0.078	0.078	0.008
	I	M	0.036	0.042	0.001	0.063	0.745	0.398	0.025	0.011	0.015	0.010	0.005	0.096
	I noS	M	0.004	0.006	0.000	0.046	0.437	0.151	0.012	0.010	0.014	0.008	0.001	0.113
	S I	M	0.215	0.181	0.426	0.775	0.529	0.673	0.486	0.958	0.682	0.570	0.082	0.630
	I&S	M	0.029	0.031	0.001	0.099	0.757	0.450	0.030	0.025	0.020	0.013	0.003	0.115
Mortality	I	F	0.165	0.044	0.111	0.079	0.491	0.317	0.305	0.280	0.011	0.002	0.051	0.019
	I noS	F	0.321	0.096	0.324	0.105	0.355	0.288	0.221	0.199	0.010	0.005	0.130	0.033
	S I	F	0.248	0.423	0.033	0.029	0.092	0.111	0.005	0.269	0.029	0.030	0.000	0.060
	I&S	F	0.140	0.048	0.045	0.029	0.368	0.231	0.100	0.249	0.005	0.001	0.006	0.010

	I	M	0.491	0.218	0.071	0.049	0.156	0.400	0.424	0.022	0.077	0.118	0.037	0.008
	I noS	M	0.318	0.138	0.093	0.045	0.064	0.254	0.178	0.005	0.072	0.081	0.007	0.001
	S I	M	0.331	0.372	0.338	0.279	0.002	0.001	0.000	0.013	0.009	0.006	0.000	0.002
	I&S	M	0.462	0.213	0.069	0.044	0.033	0.096	0.078	0.006	0.027	0.041	0.003	0.001
Lung disease	I	F	0.421	0.511	0.267	0.537	0.253	0.159	0.153	0.216	0.496	0.516	0.347	0.606
	I noS	F	0.435	0.455	0.175	0.298	0.273	0.556	0.220	0.118	0.732	0.600	0.351	0.564
	S I	F	0.185	0.283	0.291	0.331	0.219	0.262	0.189	0.616	0.329	0.412	0.432	0.442
	I&S	F	0.343	0.465	0.240	0.504	0.213	0.140	0.120	0.253	0.474	0.512	0.349	0.605
	I	M	0.102	0.040	0.017	0.254	0.069	0.102	0.591	0.361	0.241	0.346	0.574	0.952
	I noS	M	0.082	0.053	0.020	0.361	0.074	0.082	0.601	0.311	0.192	0.321	0.643	0.903
	S I	M	0.161	0.210	0.166	0.893	0.584	0.464	0.014	0.021	0.255	0.026	0.004	0.096
	I&S	M	0.073	0.031	0.012	0.363	0.085	0.111	0.330	0.179	0.215	0.207	0.303	0.909
Diabetes	I	F	0.158	0.305	0.032	0.027	0.282	0.700	0.427	0.248	0.384	0.241	0.115	0.220
	I noS	F	0.098	0.232	0.037	0.020	0.143	0.529	0.387	0.679	0.593	0.295	0.110	0.218
	S I	F	0.483	0.434	0.000	0.000	0.134	0.116	0.000	0.000	0.071	0.004	0.000	0.000
	I&S	F	0.174	0.311	0.001	0.001	0.211	0.590	0.036	0.018	0.280	0.091	0.001	0.018
	I	M	0.029	0.092	0.391	0.729	0.103	0.242	0.067	0.338	0.063	0.047	0.029	0.021
	I noS	M	0.021	0.067	0.266	0.684	0.061	0.135	0.110	0.569	0.068	0.016	0.016	0.115
	S I	M	0.794	0.723	0.009	0.029	0.419	0.506	0.185	0.621	0.115	0.039	0.329	0.061
	I&S	M	0.049	0.129	0.145	0.481	0.107	0.259	0.051	0.381	0.044	0.023	0.027	0.011
High blood pressure	I	F	0.006	0.003	0.249	0.568	0.683	0.450	0.167	0.037	0.099	0.003	0.028	0.116
	I noS	F	0.018	0.008	0.215	0.547	0.416	0.255	0.236	0.080	0.074	0.012	0.079	0.202
	S I	F	0.835	0.690	0.440	0.708	0.084	0.111	0.134	0.232	0.228	0.202	0.025	0.061
	I&S	F	0.012	0.006	0.258	0.633	0.550	0.345	0.120	0.031	0.084	0.377	0.011	0.069
	I	M	0.049	0.077	0.115	0.297	0.111	0.686	0.506	0.264	0.186	0.633	0.111	0.401
	I noS	M	0.029	0.048	0.103	0.456	0.041	0.427	0.591	0.338	0.052	0.235	0.167	0.709
	S I	M	0.327	0.257	0.201	0.085	0.035	0.168	0.260	0.082	0.091	0.041	0.335	0.253
	I&S	M	0.047	0.066	0.090	0.188	0.052	0.604	0.458	0.173	0.130	0.001	0.106	0.364
Arthritis	I	F	0.363	0.198	0.895	0.656	0.399	0.592	0.629	0.271	0.053	0.153	0.502	0.565
	I noS	F	0.234	0.107	0.974	0.697	0.352	0.388	0.459	0.676	0.015	0.040	0.199	0.668
	S I	F	0.011	0.001	0.110	0.489	0.035	0.016	0.434	0.161	0.000	0.000	0.027	0.671
	I&S	F	0.145	0.032	0.800	0.664	0.240	0.356	0.626	0.211	0.005	0.029	0.333	0.605
	I	M	0.125	0.470	0.857	0.705	0.581	0.912	0.790	0.168	0.884	0.979	0.833	0.022
	I noS	M	0.112	0.399	0.869	0.764	0.325	0.806	0.706	0.129	0.690	0.945	0.695	0.001
	S I	M	0.404	0.168	0.020	0.040	0.276	0.058	0.017	0.269	0.009	0.001	0.000	0.053
	I&S	M	0.128	0.382	0.622	0.488	0.538	0.814	0.560	0.149	0.718	0.850	0.432	0.011

Incontinence	I	F	0.195	0.256	0.186	0.231	0.540	0.758	0.241	0.824	0.185	0.321	0.081	0.109
	I noS	F	0.199	0.294	0.096	0.071	0.727	0.918	0.394	0.596	0.402	0.651	0.079	0.010
	S I	F	0.004	0.002	0.000	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	I&S	F	0.048	0.059	0.006	0.116	0.134	0.066	0.000	0.320	0.000	0.001	0.000	0.014
	I	M	0.553	0.465	0.644	0.093	0.022	0.257	0.067	0.390	0.393	0.235	0.051	0.027
	I noS	M	0.730	0.650	0.579	0.043	0.312	0.523	0.171	0.228	0.616	0.391	0.112	0.005
	S I	M	0.175	0.023	0.013	0.190	0.048	0.018	0.002	0.028	0.002	0.000	0.000	0.055
	I&S	M	0.464	0.249	0.356	0.071	0.010	0.121	0.012	0.220	0.169	0.055	0.004	0.014
Fall	I	F	0.089	0.066	n.a.	n.a.	0.444	0.544	n.a.	n.a.	0.359	0.071	n.a.	n.a.
	I noS	F	0.037	0.040	n.a.	n.a.	0.525	0.654	n.a.	n.a.	0.478	0.194	n.a.	n.a.
	S I	F	0.296	0.347	n.a.	n.a.	0.261	0.063	n.a.	n.a.	0.015	0.039	n.a.	n.a.
	I&S	F	0.081	0.064	n.a.	n.a.	0.402	0.399	n.a.	n.a.	0.205	0.038	n.a.	n.a.
	I	M	0.013	0.004	n.a.	n.a.	0.275	0.399	n.a.	n.a.	0.301	0.156	n.a.	n.a.
	I noS	M	0.005	0.001	n.a.	n.a.	0.111	0.159	n.a.	n.a.	0.073	0.030	n.a.	n.a.
	S I	M	0.009	0.049	n.a.	n.a.	0.079	0.013	n.a.	n.a.	0.015	0.047	n.a.	n.a.
	I&S	M	0.003	0.002	n.a.	n.a.	0.185	0.203	n.a.	n.a.	0.164	0.094	n.a.	n.a.
Hip fracture	I	F	0.083	0.134	n.a.	n.a.	0.569	0.392	n.a.	n.a.	0.143	0.116	n.a.	n.a.
	I noS	F	0.086	0.097	n.a.	n.a.	0.302	0.241	n.a.	n.a.	0.043	0.040	n.a.	n.a.
	S I	F	0.195	0.206	n.a.	n.a.	0.821	0.713	n.a.	n.a.	0.689	0.325	n.a.	n.a.
	I&S	F	0.064	0.108	n.a.	n.a.	0.643	0.447	n.a.	n.a.	0.170	0.110	n.a.	n.a.
	I	M	0.000	0.155	n.a.	n.a.	0.097	0.175	n.a.	n.a.	0.006	0.026	n.a.	n.a.
	I noS	M	0.036	0.384	n.a.	n.a.	0.155	0.202	n.a.	n.a.	0.006	0.034	n.a.	n.a.
	S I	M	0.011	0.032	n.a.	n.a.	0.048	0.028	n.a.	n.a.	0.203	0.382	n.a.	n.a.
	I&S	M	0.000	0.070	n.a.	n.a.	0.051	0.087	n.a.	n.a.	0.005	0.026	n.a.	n.a.
Proxy interview	I	F	0.953	0.792	0.820	0.519	0.482	0.512	0.059	0.209	0.444	0.202	0.130	0.093
	I noS	F	0.998	0.975	0.964	0.480	0.782	0.745	0.094	0.130	0.652	0.376	0.145	0.021
	S I	F	0.009	0.007	0.000	0.577	0.182	0.317	0.000	0.026	0.001	0.001	0.000	0.073
	I&S	F	0.784	0.490	0.278	0.553	0.414	0.484	0.001	0.101	0.185	0.054	0.001	0.059
	I	M	0.000	0.000	0.003	0.071	0.001	0.005	0.001	0.105	0.001	0.000	0.000	0.018
	I noS	M	0.000	0.000	0.001	0.184	0.001	0.020	0.005	0.216	0.010	0.007	0.000	0.018
	S I	M	0.000	0.000	0.000	0.000	0.004	0.001	0.000	0.000	0.000	0.221	0.000	0.000
	I&S	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000
Cognitive impairment	I	F	0.445	0.356	n.a.	n.a.	0.053	0.353	n.a.	n.a.	0.090	0.092	n.a.	n.a.
	I noS	F	0.430	0.395	n.a.	n.a.	0.013	0.234	n.a.	n.a.	0.127	0.154	n.a.	n.a.
	S I	F	0.052	0.018	n.a.	n.a.	0.000	0.000	n.a.	n.a.	0.000	0.000	n.a.	n.a.
	I&S	F	0.285	0.168	n.a.	n.a.	0.004	0.004	n.a.	n.a.	0.001	0.000	n.a.	n.a.

	I	M	0.219	0.218	n.a.	n.a.	0.034	0.244	n.a.	n.a.	0.157	0.225	n.a.	n.a.
	I noS	M	0.166	0.344	n.a.	n.a.	0.052	0.347	n.a.	n.a.	0.228	0.217	n.a.	n.a.
	S I	M	0.036	0.010	n.a.	n.a.	0.001	0.000	n.a.	n.a.	0.000	0.000	n.a.	n.a.
	I&S	M	0.109	0.076	n.a.	n.a.	0.005	0.022	n.a.	n.a.	0.003	0.000	n.a.	n.a.
Psychiatric disease	I	F	0.074	0.034	0.858	0.756	0.945	0.757	0.615	0.236	0.175	0.024	0.488	0.823
	I noS	F	0.063	0.029	0.748	0.741	0.950	0.794	0.531	0.167	0.269	0.033	0.340	0.885
	S I	F	0.041	0.004	0.000	0.000	0.737	0.208	0.028	0.027	0.299	0.196	0.000	0.000
	I&S	F	0.033	0.006	0.260	0.397	0.958	0.701	0.410	0.118	0.162	0.019	0.044	0.362
	I	M	0.066	0.204	0.506	0.790	0.006	0.038	0.357	0.043	0.088	0.013	0.260	0.747
	I noS	M	0.212	0.174	0.496	0.851	0.021	0.056	0.541	0.159	0.143	0.020	0.276	0.798
	S I	M	0.428	0.280	0.248	0.870	0.076	0.018	0.049	0.009	0.035	0.036	0.022	0.354
	I&S	M	0.071	0.183	0.450	0.856	0.003	0.013	0.224	0.011	0.048	0.006	0.144	0.730
Depression	I	F	0.199	0.224	0.224	0.438	0.732	0.207	0.292	0.708	0.068	0.175	0.293	0.827
	I noS	F	0.177	0.196	0.457	0.547	0.748	0.217	0.222	0.802	0.095	0.167	0.456	0.746
	S I	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I&S	F	0.023	0.018	0.000	0.001	0.062	0.001	0.000	0.004	0.000	0.000	0.000	0.000
	I	M	0.007	0.069	0.207	0.158	0.452	0.052	0.285	0.233	0.093	0.138	0.529	0.065
	I noS	M	0.007	0.039	0.064	0.078	0.414	0.045	0.109	0.143	0.041	0.045	0.133	0.029
	S I	M	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I&S	M	0.000	0.001	0.000	0.000	0.222	0.004	0.000	0.003	0.002	0.001	0.000	0.000
BMI	I	F	0.039	0.010	0.019	0.002	0.038	0.000	0.000	0.000	0.001	0.000	0.000	0.000
	I noS	F	0.018	0.005	0.008	0.004	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	S I	F	0.503	0.328	0.221	0.437	0.001	0.004	0.004	0.116	0.011	0.008	0.002	0.042
	I&S	F	0.046	0.010	0.015	0.002	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I noS	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	S I	M	0.931	0.828	0.260	0.388	0.316	0.182	0.006	0.102	0.487	0.007	0.004	0.147
	I&S	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Smoke now	I	F	0.085	0.039	0.065	0.171	0.421	0.231	0.073	0.004	0.031	0.049	0.015	0.038
	I noS	F	0.062	0.061	0.031	0.161	0.655	0.343	0.228	0.113	0.046	0.031	0.007	0.261
	S I	F	0.140	0.046	0.000	0.000	0.309	0.084	0.000	0.000	0.172	0.166	0.000	0.000
	I&S	F	0.060	0.018	0.010	0.006	0.396	0.159	0.005	0.000	0.024	0.038	0.000	0.000
	I	M	0.000	0.904	0.103	0.351	0.577	0.087	0.700	0.170	0.221	0.415	0.087	0.003
	I noS	M	0.679	0.000	0.173	0.564	0.016	0.076	0.543	0.315	0.374	0.382	0.055	0.001
	S I	M	0.459	0.053	0.002	0.016	0.400	0.302	0.000	0.003	0.305	0.079	0.000	0.000
	I&S	M	0.000	0.794	0.019	0.162	0.569	0.081	0.287	0.046	0.208	0.325	0.002	0.000

ADL impairment	I	F	0.046	0.101	0.066	0.058	0.193	0.210	0.024	0.068	0.000	0.000	0.000	0.051
	I noS	F	0.031	0.072	0.079	0.020	0.181	0.244	0.036	0.030	0.000	0.000	0.000	0.028
	S I	F	0.238	0.295	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I&S	F	0.039	0.093	0.002	0.004	0.042	0.012	0.000	0.000	0.000	0.000	0.000	0.000
	I	M	0.026	0.011	0.003	0.059	0.397	0.414	0.675	0.254	0.018	0.180	0.324	0.423
	I noS	M	0.305	0.264	0.038	0.019	0.468	0.688	0.850	0.443	0.264	0.386	0.381	0.208
	S I	M	0.453	0.197	0.000	0.000	0.639	0.034	0.000	0.000	0.081	0.000	0.000	0.000
	I&S	M	0.029	0.009	0.000	0.003	0.433	0.273	0.003	0.001	0.011	0.036	0.000	0.004
IADL impairment	I	F	0.467	0.602	0.577	0.546	0.108	0.008	0.006	0.344	0.442	0.243	0.131	0.052
	I noS	F	0.312	0.395	0.507	0.600	0.106	0.021	0.004	0.154	0.337	0.103	0.080	0.074
	S I	F	0.114	0.056	0.001	0.007	0.110	0.030	0.000	0.003	0.053	0.011	0.000	0.003
	I&S	F	0.364	0.443	0.199	0.261	0.075	0.003	0.000	0.081	0.335	0.129	0.003	0.014
	I	M	0.009	0.166	0.097	0.201	0.137	0.105	0.005	0.004	0.053	0.007	0.014	0.010
	I noS	M	0.003	0.092	0.071	0.251	0.084	0.082	0.021	0.015	0.010	0.008	0.005	0.014
	S I	M	0.023	0.027	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I&S	M	0.003	0.078	0.000	0.006	0.031	0.007	0.000	0.000	0.002	0.000	0.000	0.000
Self-rated health	I	F	0.004	0.001	0.003	0.084	0.000	0.004	0.004	0.011	0.000	0.000	0.000	0.034
	I noS	F	0.024	0.003	0.014	0.056	0.000	0.003	0.005	0.005	0.000	0.000	0.000	0.009
	S I	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I&S	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I	M	0.052	0.017	0.020	0.216	0.039	0.016	0.021	0.063	0.004	0.001	0.001	0.011
	I noS	M	0.019	0.009	0.016	0.468	0.009	0.010	0.004	0.037	0.006	0.003	0.001	0.018
	S I	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	I&S	M	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Results are for white females (F) and males (M). Abbreviations of tests are as follows: I = Unconditional invariance; I|noS = Invariance, conditional on non-causality; S|I = Non-causality, conditional on invariance; I&S = Joint invariance and non-causality.