This is a repository copy of *The Impact of Health on Labor Supply Near Retirement*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/194012/

Version: Published Version

Article:

The Impact of Health on Labor Supply Near Retirement

Richard Blundell
University College London

Jack Britton
Institute for Fiscal Studies

Monica Costa Dias
Institute for Fiscal Studies

Eric French
University College London[*]

October 2020

Abstract

Estimates of how health affects employment vary considerably. We assess how different methods and health measures impact estimates of the impact of health on employment using a unified framework for the US and England. We find that subjective and objective health measures, and subjective measures instrumented by objective measures produce similar estimates when using sufficiently rich objective measures. Moreover, a single health index can capture the relevant health variation for employment. Health deterioration explains up to 15% of the decline in employment between ages 50 and 70. Effects are larger for the US than England, and for the low educated.

I10, J24, J26, E24

Keywords: Health, Cognition, Labor Supply, Retirement

I   Introduction

Despite the growing literature and the increasing availability of rich data, there is still no consensus about the importance of health for employment. The existing literature has developed many empirical approaches and applied them to different datasets collected in different contexts. This naturally led to estimates of the effects of health on employment that differ significantly from study to study. Currie and Madrian (1999), O'Donnell et al. (2015) and French and Jones (2016) review the empirical evidence and advance some potential explanations for the discrepancies between estimates. Most of these relate to the measurement and modeling of health.[i]

Ideally one would like to have a composite index of health representing 'working capacity' or 'health stock' – a comprehensive description of health status that could be used in a variety of contexts and facilitate comparisons across studies. The difficulty, of course, resides on the fact that such an index is not readily observable. This has led to a proliferation of different methods to proxy it. For instance, some applications adopt a multi-dimensional description of health, with many variables affecting employment in a flexible way; other applications rely on a constructed health index that is then related to employment. The type of information used to describe health also varies across studies. Some use 'objective' indicators, which unambiguously describe specific health conditions (such as arthritis), while others use 'subjective' accounts of self-reported health to obtain a comprehensive measure of health status. Furthermore, there is no agreement about which specific objective and subjective health variables should be used. Moreover, various modeling strategies have also been adopted, often resulting in different estimates of the effect of health. For instance, studies using cross-sectional data tend to focus on the overall impact of health, while longitudinal data can be used to estimate the impact of changes in health.

Despite the important differences, there is still little systematic research assessing the relative merits of the various methods. In this study, we aim to fill this gap by addressing the following questions. Is the choice of health measure important for measuring its impact on employment? How should the health measures becoming available in survey data be combined into a health index? Is a single health measure sufficient to capture the impact of health on employment, or is it important to allow for multiple measures? Are cross sectional methods appropriate, or is it necessary to consider individual heterogeneity by accounting for initial conditions?

To answer these questions, we revisit many of the approaches proposed in the literature within a unified framework. We produce a set of estimates that can be compared across specifications, and contrast the resulting estimates using formal statistical tests, relating their differences to the underlying measurement and modeling choices. Specifically, we compare estimates of health effects obtained by using either subjective measures or objective measures. We deal with various sources of measurement error, including justification bias, by combining the two sets of health variables and using the objective measures as instruments for the subjective measures. We recognise that some of the objective health measures may suffer from the same sources of justification bias as the subjective health measures, and test for this by restricting the set of instruments to the most serious conditions that require urgent medical attention. We use principal components and factor analysis to construct a parsimonious single health index that summarises information from multiple health measures. An index of the common variation across these variables is likely to be a better summary of health status than any of the original measures taken individually, and is likely to be less sensitive to measurement error. We enlarge our empirical model to include cognition, a dimension that is not typically considered in other studies but that is closely intertwined with health and may capture a finer detail of how poor health impairs work.

Our empirical analysis is based on two large longitudinal surveys of older people, the US Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA). These are high-quality longitudinal datasets that include many different measures of health, all key requisites to support the replication of the alternative measures and models of health and employment used in past studies. Moreover, their very similar structures and information supports the use of harmonized measures and estimation procedures in producing comparable estimates for the two countries.

Our key findings are as follows. First, we find that objective and subjective health measures deliver similar estimates if a sufficiently large set of objective measures is used; controlling for only a limited number of health conditions, however, may reduce the estimated impact of health on employment by two thirds. Second, we find that a single health index, while sometimes rejected from a statistical standpoint, produces estimates of the effect of health on employment that are similar to those obtained using multiple health indexes. Third, using objective measures to instrument for subjective measures also produces similar, although slightly larger estimates. Fourth, we find that properly accounting for heterogeneity in background characteristics by controlling for initial conditions is a more important modeling issue than the choice of the health measure. Fifth, although cognition is significantly related to employment, we find that it has little added explanatory power once we also control for health, suggesting that cognition is not a key driver of employment at these ages.

For direct comparison across groups, countries and methods, we calculate the share of the decline in employment between ages 50 and 70 that can be explained by declines in health. Overall we find that, depending on country, gender and education, declines in health explain between 3% and 15% of the decline in employment. These effects are larger for high school dropouts and tend to decline with education. They are also larger in US than in England, generally by a factor of two to three. We estimate that the majority of the differences across

countries is driven by the stronger effect of health on employment in the US, rather than by differential declines in health or employment. However, the key findings we outline above are consistent across the two countries.

The rest of the paper is outlined as follows. Section II provides an overview of the literature investigating the impact of health on labor supply. Section III outlines the methods we use to measure health and cognition, and develops a unifying framework under which the most commonly used models of health and employment can be compared. Section IV describes the ELSA and HRS datasets and our constructed measures of health and cognition. Section V presents our main estimates and examines the sources of differences between the US and England. Section VI presents a simple dynamic structural model of employment and retirement with health, and uses the model to discuss the various mechanisms through which health affects employment and our empirical strategy. Section VII concludes.

II   Literature

This paper brings together several strands in the literature on health and employment. First, it relates to the large literature aiming to quantify the impact of health on employment and to establish the relative merits of subjective health measures, objective health measures and subjective measures instrumented by objective measures in estimating this effect. Concerns about various sources of bias afflicting estimates using each of these measures have impeded comparisons across studies and precluded the emergence of a clear picture on the importance of health effects. On their own, objective indicators describe diagnosed health conditions but relate only to a subset of the relevant conditions and miss severity information, hence providing an incomplete view of health. In turn, subjective indicators offer a comprehensive view of health status, but are often crude categorical measures of health and are particularly vulnerable to reporting error. However, subjective measures instrumented by objective ones are immune to the measurement issues afflicting each set of measures taken independently if these are

unrelated, and can therefore be used to benchmark estimates using only one type of health measure. We use the three approaches to assess and quantify how measurement error, justification bias and limited health information bias estimates of the impact of health on employment.

Early research suggests that subjective measures produce significantly larger estimates of the impact of health on employment than objective measures. For example, Bound (1991) found differences of nearly one order of magnitude when using future mortality as an objective health measure. However, estimates relying exclusively on objective variables tend to use more detailed health information than Bound (1991) did. For instance, Bartel and Taubman (1979) uses variables describing heart disease, psychiatric conditions, arthritis and asthma; more recent work using the Health and Retirement Survey (HRS) enlarges this list (e.g. Smith, 2004). We add to this literature by including more objective variables and by showing how adding information on health conditions changes the estimated effect. Consistent with past results, we find that limiting the number of objective measures produces estimates that are significantly smaller than those obtained using subjective measures. However, these differences vanish once a sufficiently large number of objective measures is used.

In turn, there are widespread concerns that estimates using subjective measures are biased up due to justification bias, whereby non-working individuals tend to report lower levels of health partly to justify their work status (e.g. Butler et al., 1987). The extent of justification bias has been heavily studied, with mixed results. Benitez-Silva et al. (2004) cannot reject the hypothesis that self-reported disability is an unbiased measure of true disability, while Kreider and Pepper (2007) find that non-workers tend to over-report disability rates. However, subjective measures are also subject to other forms of reporting error, particularly as they are often relatively crude measures. Such measurement error may lead to attenuation bias in the estimates of health effects, which will at least partly counteract the effect of justification bias.

6

Studies of measurement error in subjective measures show that it is not negligible. For instance, Crossley and Kennedy (2002) find that 28% of all respondents change their reported health status when being asked the same self-assessed health question twice in the same interview (French, 2005, shows similar evidence of misreporting).

Stern (1989) suggests using objective measures to instrument for subjective measures. Bound (1991) shows that this procedure produces estimates that are close to those using subjective measures, suggesting that measurement error and justification bias in subjective measures roughly offset. Dwyer and Mitchell (1999), McGarry (2004), and Giustinelli and Shapiro (2018) circumvent concerns of justification bias by examining the relationship between health and *expected* retirement. Similarly, Pamela and Shapiro (2018) use responses to hypothetical questions about people's retirement decisions given different hypothetical health levels. Their approach is to focus on those who have not yet retired and who, therefore, do not need to justify retirement on bad health. They find strong links between subjective health measures and expected retirement. We contrast estimates using subjective measures, objective measures, and objective measures instrumenting for subjective measures, and find that all three approaches produce surprisingly similar estimates when using the full set of objective measures available in the HRS and ELSA.

Second, this paper also connects to the literature contrasting cross-sectional and panel data methods in estimating the impact of health. It has been noticed that cross-sectional estimates are vulnerable to reverse causality and simultaneity, both leading to upward bias. For instance, it is conceivable that higher incomes cause better health. The Grossman (1972) model implies that those with higher income may be able to purchase better nutrition and health care, improving later health outcomes. On the other hand, the simultaneous determination of health and employment could result from common (unobserved) drivers of both outcomes. For instance, it may be the case that high-income parents invest more in both the health and the

education of their children, leading to better health and income outcomes later in life. In line with this view, Case et al. (2002) show that child health is positively related to household income and, most importantly, that this relationship becomes stronger over time, as the child ages.

Panel data methods offer the tools to deal with the confounding effects of reverse causality and simultaneity bias. Smith (2004), Blau and Gilleskie (2001) and Gilleskie and Hoffman (2014) emphasize the difference between panel and cross sectional methods for the purpose of estimating health effects, and we revisit this issue. We find that including a full set of initial conditions and focusing on estimating the impact of changes in health on employment reduces the magnitude of the health coefficients by half. These findings are consistent with non-negligible bias induced by reverse causality and simultaneity.

The final strand of the literature to which this paper relates is that assessing the ability of parsimonious representations of health to capture the relevant finer detail present in multiple measures. A parsimonious representation of health is especially valuable in contexts where high-dimensional problems are impractical, such as when estimating complex models. But whether the single index is a sufficiently detailed representation of health remains an open question. We show that a single health index captures well the variation in health that matters for employment. To the best of our knowledge, we are the first to test the single index assumption in this way. The closest example in the literature is Blau and Gilleskie (2001), who argue that 'no single measure of health is adequate to explain labor force transitions of older men'. They draw this conclusion from a series of estimates that add, sequentially, more subjective and objective measures in the HRS. We obtain similar results to Blau and Gilleskie (2001) when gradually adding more objective and subjective health variables in our employment estimates. But we find that a single measure combining several subjective health

variables through principal components analysis is sufficient to capture the overall impact of health on employment.

III   Methods for Estimating the effect of health and cognition on employment

Despite the growing literature on the effect of health on employment, there is still no agreement on its magnitude. The lack of consensus may be partly due to the variety of empirical approaches and datasets that have been used to measure these effects. A key source of differences relates to how health is measured. Ideally one would like a summary measure of health linked to work capacity, but such a measure is not readily observed in the data. Current datasets do not include all the health variables that affect work capacity, and those that are included may suffer from measurement error and justification bias to different degrees. Alternative estimation approaches deal differently with these problems as we discuss below.

Here we bring together these approaches under a common unifying framework to contrast their predictions and assess the validity of their underlying assumptions. Specifically, we address the following issues: (1) how should we expect estimates of the effect of health on employment to differ when using objective versus subjective measures? (2) how should using objective health measures to instrument for subjective measures affect the estimates? (3) is a single health index sufficient, or should multiple health indexes be used to capture the effect of health on employment? We show how to use multiple objective and subjective measures to answer these questions.

Our analysis is based on a simple empirical model of employment, for which we consider two alternative but similar specifications. The first uses a linear probability framework. For individual i at time t:

(1) $$Y_{it} = \theta_0 + \theta_H H_{it} + \theta_X X_{it} + e_{it}$$

where Y is a binary indicator of employment, H^* is health status, with the superscript highlighting that health is not directly observed in data, and X are other drivers of employment, which we discuss in detail below. The second specification assumes that Y is determined by the, latent index Y^* as follows:

$$Y_{it}^* = \theta_0 + \theta_H H_{it}^* + \theta_X X_{it} + e_{it}$$

$$Y_{it} = \mathbf{1}\{Y_{it}^* > 0\} \tag{2}$$

In this case, we will assume that e_{it} is normally distributed and thus estimate the model using a probit.

These employment equations are derived from a structural model of life-cycle labour supply and health, which we present and discuss in Section VI. The structural model provides an interpretation for the parameter of interest in this study, which is θ_H. Expressions (24) and (25) in Section VI, which are derived directly from the economic model of behaviour, demonstrate the many ways in which health affects employment that are subsumed into θ_H. These include the impact of health on the utility cost of work, pay from work, entitlement to benefits such as those for disability, and expectations about future work, pay and lifespan given the persistent nature of health. Our empirical analysis will not allow us to disentangle these various mechanisms. Instead, the focus of this study is on how to estimate the overall effect of health on employment through all of the above channels (θ_H) in ways that are robust to measurement error in health and to biases from self-justification or other sources.

The structural model also guides our choice of the other covariates in the regression equations, which we denote by X. In this paper we are not interested on the value of their related parameters, but it is nevertheless important that we control for the right set of covariates in order to understand how to interpret estimates of θ_H. For instance, the structural analysis in Section VI demonstrates that it is important to control for age and time in order to capture how

preferences for work and the monetary incentives to so (including pay for work and benefits) change around the age of retirement and differentially for different generations. Therefore, X includes time dummies and a second order polynomial in age. Our structural analysis also reveals the need to control for initial conditions in health and employment, which are meant to capture permanent heterogeneity in preferences, productivity, and health. If these initial conditions are not included, estimates of θ_H would be confounded by unobserved factors driving both employment and health. One issue that the structural model shows is that it is the initial employment index (Y^*) capturing the propensity to work that ought to be accounted for in the initial condition. That index, however, is not observed; what is observed instead is employment status (Y). Using the structural model, we characterise what governs the latent index Y^*, and complete the initial condition for employment with those for its other determinants in the initial period. These include work experience, wealth, marital status and the fixed health traits that we capture by health status during childhood. Conditionally on this rich set of covariates, we then assume that the health status H^* is independent of the unexplained driver of employment, e.

In what follows, we discuss the measurement of health and the identification and estimation of the parameter of interest, θ_H. In discussing the potential bias in alternative estimation procedures we will, for simplicity, focus on the linear employment equation specified in equation 1. All results also hold for the probit specification in equation 2.

A   Measuring health using objective measures

The health stock can be formalized by a combination of all health conditions (and combinations of conditions) that limit work, h_k^o for k = 1,…,K. These are typically labeled 'objective' health measures because they represent medical health conditions that can be unambiguously identified; indeed, some surveys report only conditions that have been medically diagnosed and for which the respondent receives treatment.

Assuming a linear functional form, we write

(3) $$H_{it}{}^* = \sum_{k=1}^{K} \alpha_k\, h_{kit}^{o}$$

and this expression can be replaced in equation (1) to yield

(4) $$Y_{it} = \theta_0 + \sum_{k=1}^{K} \tilde{\theta}_{Hk}\, h_{kit}^{o} + \theta_X X_{it} + e_{it}$$

where $\quad \tilde{\theta}_{Hk} = \theta_H \alpha_k$

In practice, the simple specification in equation (4) is sensitive to potentially serious measurement problems for four reasons. First, the number of observed conditions K^o is smaller than the total number of health conditions K since one can only ever observe a limited subset of the relevant medical conditions. This is true even if one has full access to medical records, as only diagnosable conditions under current technology can be observed. Health status can be decomposed into observed and unobserved objective conditions:

(5) $$H_{it}{}^* = \sum_{k=1}^{K^o} \alpha_k\, h_{kit}^{o} + q_{it}$$

where q summarises the contribution to health status of the $K - K^o$ unobserved conditions. Consequently, the effect of health can only partly be determined. Second, not all health conditions are equally important for overall health and thus employment, a fact that is expressed by the multiple parameters $\tilde{\theta}_{Hk}$. While some conditions may be so debilitating as to completely impair work (like strokes) others may have more limited consequences for work capacity (like diabetes). Hence, the magnitude of the estimated impact will depend critically on exactly which conditions are accounted for. Third, estimates of the impact of specific observed conditions may be biased if unobserved conditions are related to observed ones. And fourth, most health measures only describe whether respondents suffer from certain conditions, not the severity of those conditions. This is a key source of measurement error biasing the estimated effects, potentially towards zero.

To put it more formally, consider the linear regression model of employment in equation 1 and assume that the true health stock H^* is a combination of two conditions, (h_1^o, h_2^o). For this discussion we also ignore the correlation between health and the X variables. We normalize the variance of the objective measures to equal that of H^*,[ii] and ensure that all variables are ordered in the same direction (say, higher values for better health) so that $(\alpha_1, \alpha_2) \in [0,1]$[ii]. Suppose that h_1^o is observed and measured without error, but h_2^o is unobserved. In such case, the OLS estimator of θ_H yields

$$\text{plim } \hat{\theta}_H^o = \frac{\text{Cov}(Y, h_1^o)}{\text{Var}(h_1^o)}$$

$$= \frac{\text{Cov}(\theta_0 + \tilde{\theta}_{H1} h_1^o + \tilde{\theta}_{H2} h_2^o + \theta_X X + e, h_1^o)}{\text{Var}(h_1^o)}$$

$$= \theta_H \alpha_1 + \theta_H \alpha_2 \frac{\text{Cov}(h_1^o, h_2^o)}{\text{Var}(H^*)}$$

If $\text{Cov}(h_1^o, h_2^o) = 0$ then $\text{plim } \hat{\theta}_H^o = \theta_H \alpha_1$ and will thus identify the effect of the first health condition, which is smaller than the impact of the global health measure (θ_H) under the assumptions stated above. Moreover, had one observed h_2^o instead of h_1^o, a different impact would be identified (specifically, $\theta_H \alpha_2$).

In the likely case where the two health condition measures are positively correlated (with a second health condition being more prevalent among those who already suffer from the first health condition), then the estimated effect of health will be closer to the true overall effect (hence less biased) than under the case where they are uncorrelated. A prediction based on model estimates of how much changes in health status drives employment (as described below in Section III.G) will still be biased towards zero for two reasons: first, the likely attenuation bias in the estimated coefficient, and second, the failure to account for all the relevant variation in health in the presence of missing variables.

Applications that use objective health measures often combine information from numerous health conditions. This may attenuate the estimation bias but will generally not eliminate it. With many health measures, the formula for the asymptotic limits described above becomes more complex, although the key insight is the same: the index will understate the true causal effect of health on employment because it does not capture all relevant variation in health, and the extent of the bias depends on how strongly correlated the omitted variables are with the observed ones. In fact, using any linear combination of the observed health measures (such as the first principal component of the objective measures) will understate the true causal effect. The lack of detailed medical data on the severity of a condition can be viewed as a specific case of missing variables and will, as in the general omitted variable case, lead to attenuation bias.

In the empirical application, we use the complete set of medically diagnosed conditions (for which the respondent is getting treatment) common to the two datasets. These amount to ten objective measures in total. We have produced a parallel set of results by augmenting the set of objective measures with observed variables measuring Activities of Daily Living (ADL), which are meant to capture general levels of health that may limit work. Our results are not sensitive to this choice.[iii]

B   Measuring health using objective measures

Although we cannot observe H^* directly, we do observe 'subjective' measures h_k^s for k = 1,…Kˢ. These are self-reported health measures that describe overall health status and provide an alternative to using objective measures to describe heath. The literature has interpreted the subjective measures as noisy measures of a single latent health stock H^*. Thus, while the different objective measures describe different subcomponents of the health stock (as shown in equation (3)), the subjective measures are overall (noisy) measures of the single latent health stock. This idea can be formalized by the following set of measurement equations, which relate

the observable subjective health indicators h_k^s to the unobservable latent health index H^*:

(6) $$h_{kit}^s = \beta_k H_{it}^* + u_{kit} \text{ for } k = 1, \ldots K^s$$

where u_k represents the measurement error in observed health variable k.

In practice, studies that model health as a latent variable typically use a single indicator of health (Bound et al., 1999; Bound et al., 2010; Disney et al., 2006). Instead, we use all the subjective measures of health that are contained in both the HRS and ELSA surveys, which total three, and extract a health index using Principal Component Analysis.[iv] This is a natural approach if one wants to summarise the common information in many subjective measures, each being a noisy measure of the same latent health variable.[v] It turns out that the results are not sensitive to the procedure used to extract the variation from the subjective measures; we show only results using Principal Components Analysis in the main text (see Online Appendix Section 4.2 for some results using Factor Analysis).

Let H^S be the subjective health index constructed using the subjective health measures. The single index is a parsimonious approach that can be used in a variety of contexts; it is particularly useful when keeping the number of health variables low is paramount, such as for estimation of structural models of health. Moreover, the use of common variation across many subjective health measures (using approaches such as factor analysis or principal components analysis) helps mitigate the importance of measurement error if the noise across different variables is independent.

However, measurement error is unlikely to be completely eliminated by the use of many measures in constructing the health index. In particular, justification bias affecting all underlying subjective measures implies that measurement error is not classical. So we write

(7) $$H_{it}^S = H_{it}^* + v_{it}$$

If the unobserved component of employment (e) and the measurement error (v) are uncorrelated, estimates of the health effect θ_H will be biased towards zero. In the more likely

event that (e, v) are positively related – those not working tend to report lower levels of health partly to justify their working status – the direction of the overall bias is ambiguous. Indeed, the OLS estimator of θ_H in equation 1 using H^S to proxy H^* has asymptotic limit:

$$(8) \qquad \text{plim } \hat{\theta}_H^s = \frac{\theta_H \text{Var}(H^*) + \text{Cov}(e,v)}{\text{Var}(H^*) + \text{Var}(v)}$$

which may be greater or smaller than the parameter of interest θ_H depending on the sign and relative size of $\text{Cov}(e, v)$. O'Donnell et al. (2015) suggest that justification bias dominates and $\text{Cov}(e, v) > 0$, resulting in an upward biased estimate of θ_H. However, Stern (1989) and Dwyer and Mitchell (1999) do not find that justification bias dominates.

C   Using instrumental variables to deal with measurement error and justification bias

Thus far we have seen that approaches using exclusively objective measures suffer from omitted variable bias and are likely to produce estimates of the impact of health that are downward biased. Approaches using only subjective measures suffer both from measurement error and justification bias, leading to estimates that could be either upward or downward biased. One way of dealing with the biases afflicting estimates based on subjective health measures is to use instrumental variables. We have many potential instruments to choose from if measurement error and justification bias in the subjective measures are independent from objective health conditions, namely the entire set of objective health measures.

It is straightforward to see that any subset of the objective health measures can be used to instrument the subjective index. For simplicity, consider the case where we only have one objective measure (indexed k) and use it to instrument the subjective health index. The first stage regresses H^S on h_k^o and the estimated coefficient (call it $\hat{\eta}$) converges in probability to

$$\text{plim } \hat{\eta} = \eta = \frac{\text{Cov}(H^S, h_k^o)}{\text{Var}(h_k^o)}$$

$$= \frac{\text{Cov}(H^*, h_k^o)}{\text{Var}(H^*)}$$

16

$$= \frac{\alpha_k \text{Var}(H^*) + \sum_{l \neq k} \alpha_l \text{Cov}(h_l^o, h_k^o)}{\text{Var}(H^*)}$$

Recall that H^* is a combination of all objective health conditions (as described in equation (3)), each of which has been standardized to have a variance equal to that of H^*.

The predicted value of H^S is, therefore, $\hat{\eta} h_k^o$. The second stage instrumental variables estimate using the linear employment equation 1 is

$$\text{plim } \hat{\theta}_H^{IV} = \frac{\text{Cov}(Y, \eta h_k^o)}{\text{Var}(h_k^o)}$$

$$= \theta_H \frac{\text{Cov}(H^*, h_k^o)}{\eta \text{Var}(H^*)} = \theta_H$$

Under the IV exclusion restrictions, we can assess the importance of biases confounding estimates of θ_H based on objective measures (due to omitted variables) and based on subjective measures (due to measurement error and justification bias). We do this by comparing IV estimates to those obtained using only objective or subjective health measures.

It is straightforward to show that the IV approach is valid even if there is measurement error in the objective measures, so long as that measurement error is orthogonal to that affecting the subjective measures. In particular, this assumption requires that the justification bias generally associated with subjective measures does not permeate into responses to the survey questions on objective health measures. We discuss the plausibility of this assumption in the Online Appendix.

D   Tests of the single index assumption

We now turn to discuss the plausibility of the single index assumption. The 'single index assumption' states that there exists an index of multiple measures of self-reported health status H^S, constructed as a composite measure of the subjective health variables, that contains all relevant health information for employment. Under this assumption, the objective measures

impact employment only through their impact on H^S. This is a restriction on model (1) in which the latent measure of health (H^*) can be a function of multiple health conditions with varying implications for work capacity as described in equations (3) and (4). We use this restriction to derive a specification test below. Notice that measurement error and justification bias are not ruled out by this assumption. Indeed, we do allow for both sources of noise in H^S, as described in equation (7). The single index assumption imposes that any measurement error in H^S (v in equation (7)) is independent of H^*:

$$v_{it} \perp H_{it}^*$$

The single index assumption underpins much of the empirical work on the impact of health on labor supply. In particular, it is critical in contexts where dealing with multiple health dimensions is impractical, such as in large structural models. We now use our methods to assess the validity of this assumption using data that is now becoming widely available in developed countries. To the best of our knowledge, this has not been done before.

First, we use our subjective measures. Under the single index assumption, all subjective measures of health are noisy measures of the same concept. Thus, each individual measure should have little predictive value for employment above and beyond a summary measure of all subjective variables. We test this assumption by including the Second and Third Principal Components of health in the employment model, in addition to the First Principal Component. Formally, we test the explanatory power of the added principal components.[vi]

Second, we use the objective measures to assess the single index assumption. One simple point is that the single index assumption implies that the effect of health estimated using the index should not be smaller than that estimated using objective measures. This is because a correctly specified health index should capture all relevant health information for employment, while objective measures can only capture part of the relevant variation (as explained above).

We therefore compare the magnitude of the health effects based on the single subjective health index and the full set of objective measures.

A slightly subtler point is that the IV approach with multiple instruments provides the means to test the validity of the single index assumption using a Sargan overidentification test (Hansen, 1982). The intuition is simple: if the single index assumption is valid, all the objective measures (the instruments) should affect labor supply only through the subjective health index. For this reason, the IV residuals e^{IV} should not be correlated with the instruments. With ten objective measures, we have nine overidentification conditions.

In practice, we implement the test following the suggestion in Davidson and MacKinnon (2003). For the linear probability regression model in equation 1, we construct the IV residuals:

$$(9) \qquad \hat{e}_{it}^{IV} = Y_{it} - \hat{\theta}_{it}^{IV} - \hat{\theta}_{H}^{IV} H_{it}^{s} - \hat{\theta}_{X}^{IV} X_{it}$$

Under the single index assumption, we know that:

$$(10) \qquad E[\hat{e}_{it}^{IV} h_{kit}^{o} | X_{it}] = 0 \quad \text{for } k = 1,..., K^{o}$$

So we regress the residual on all health objective measures and the exogenous variables *X*, and calculate the F-statistic associated with the hypothesis that all health coefficients are jointly equal to zero. For the latent index model of employment in equation 2 we use the over-identification test developed in Lee (1992), based on the minimum distance estimator proposed in Newey (1987) (see also Rivers and Vuong, 1988).[vii]

E   Health measures and the estimation of the effects of other determinants of employment

The focus of this paper is on obtaining consistent estimates of the impact of health on employment θ_H. Here we discuss how our various approaches to measuring health affect estimates of the effects of other drivers of employment θ_X. We point out that instrumenting

subjective health using objective measures of health can deliver consistent estimates of θ_H but will not in general deliver consistent estimates of θ_X a result highlighted in Bound (1991).

In the context of a structural model such as that discussed in Section VI of this paper, health affects both employment and, throughout life, other choices and outcomes including savings towards retirement, offered wages, and other financial incentives to retire (see also Gilleskie et al., 2017a). In the statistical model of employment in equation (1), which can be derived from that structural model, these other economic variables are contained in *X*. However, since health status is not observed, the statistical model of employment needs to be completed with a description of how health is measured. In this paper we consider two alternative proxy measures, based on objective and subjective health measures as expressed in equations (5) and (7), respectively. The error in these measures may be correlated with the other drivers of employment (*X*) in ways that bias the estimates of their effects. Moreover, since the nature of that error differs across health measures, so will its consequences for the estimation of the effects of *X* on employment.

To be more specific, consider first the use of an incomplete set of objective health measures as described in Section III.A. If the omitted health conditions (h_2^o) in the notation of that section) are correlated with the covariates *X* in the employment equation 1, the resulting estimates of θ_X will be biased. To focus ideas, consider estimates of the effect of age on employment. In our specification, the age coefficient is particularly interesting as it summarises the joint roles of changing preferences and monetary incentives to retire in driving employment of older workers. We expect it to be negative, as older people are increasingly less likely to work. If health deteriorates faster later in life and in ways that are not fully captured by observed objective health h_1^o, then the age coefficient would partly encapsulate the effects of the deterioration of health along the unobserved dimensions h_2^o. In this case, estimates of the age

effects would be downward biased, away from zero, which means that one would overplay the effect of age on retirement.

The consequence of using a subjective health index for the direction and magnitude of the bias of the age effects on employment near retirement is more ambiguous. To be concrete, suppose that justification bias dominates other sources of measurement error, so that $\text{Cov}(e, v) > 0$ in equation (8); this will result in upward biased estimates of the impact of health on employment (θ_H). The mismeasurement of H^* can affect estimates of the age effects in two ways that may partly cancel out. First, the bias in θ_H leads to an over-prediction of the role of health deterioration with age in driving employment. In our regression model, this would be partly compensated by an age effect biased towards zero. That would be the only source of bias if age is independent of the justification error, but not otherwise. The second source of bias arises precisely if the measurement error in health status is correlated with age. For instance, one could think that the importance of justification bias fades with age if old age is widely accepted as a valid reason for not working, in which case younger workers under-report their health more strongly than older workers. This would bias the estimate of the age coefficient downwards or away from zero partly to compensate for the fact that subjective health under-predicts the true pace of health deterioration with age. The ultimate direction of the bias in θ_X would, in this case, be undetermined a priori. If, on the contrary, justification error becomes more important with age as more workers stop working, then instead subjective health would over-predict the true pace of health deterioration with age and the two sources of bias would push the age coefficient in the same direction, towards zero.

Finally we note that instrumenting subjective health using objective health would remedy bias from the first source by producing consistent estimates of θ_H. However, it will have no impact on the second source of bias given that subjective health rather than actual health is used in the statistical model.

F Cognition

Cognition is not only a determinant of productivity in work, it may also affect work capacity in a way that is not otherwise observed in objective and subjective health variables. It may, therefore, be a critical driver of labor supply and we are interested in determining its effect. We therefore enlarge our model to control for cognition. We observe several measures of cognition, described in Section IV.D below. These are test scores, measured by the interviewer, and thus not subject to the sources of bias that may afflict health measures. Yet, our cognition measures will provide only an incomplete representation of cognitive ability, implying our estimates of the cognition effects may be biased towards zero. Denoting the latent cognition index by C^*, the extended model is

$$(11) \qquad Y_{it} = \theta_0 + \theta_{\mathrm{H}} H_{it} * + \theta_{\mathrm{C}} C_{it}^* + \theta_{\mathrm{X}} X_{it} + e_{it}$$

As in the case of health, we construct a parsimonious representation of cognitive ability under the single index assumption by summarising the cognition variables in a single index using Principal Component Analysis.[viii] When using this extended model, we supplement the initial conditions in X with cognition measured when each individual is first observed.

G Comparable measure of the impact of health and cognition

To facilitate the comparability of results across the various specifications, we construct a global measure of the impact of health or cognition by predicting their cumulative impact on employment over the 20-year period that span from 50 to 70 years of age. The parameter we calculate is

$$(12) \qquad \hat{\delta}_{\mathrm{M}} = \frac{\Theta_{\mathrm{M}}(\bar{M}_{70} - \bar{M}_{50})}{\bar{Y}_{70} - \bar{Y}_{50}}$$

where the upper bar represents average predictions from a fixed effects regressions of measures M (for health and cognition) and Y (for employment) on age. Hence, $\bar{X}_{70} - \bar{X}_{50}$ (for $X = Y, M$) is simply the average change in measure X that individuals experience between the ages of 50

and 70. The fixed effects net out differences across cohorts and attrition in the panel that could confound our estimates of individual-level decline in health, cognition or employment.

In measuring changes in health and cognition as workers get older, we rely on the exact same measures that were used to estimate each model. So the change in health or cognition that we consider to calculate $\hat{\delta}$ will depend on which specific measure was used in estimating Θ. For instance, we use changes in subjective health and in instrumented subjective health to quantify the impacts implied by estimates based on the respective measure. If the subjective health is afflicted by justification error that varies with age then that age dependence will be reflected on our measure of health deterioration in the 20 years from age 50 based on the subjective index *but not* on its instrumented counterpart.

When using various measures of health and cognition together in the same regression model – such as, for instance, when estimating a model of employment on objective health measures – we use changes in each measure to calculate the single impact parameter

(13) $$\hat{\delta} = \sum_j \frac{\Theta_j(\bar{M}_{j,70} - \bar{M}_{j,50})}{\bar{Y}_{70} - \bar{Y}_{50}}$$

where j indexes the various health and cognition measures included in the employment regression model. Here again Θ_j is the marginal effect of health or cognition measure j evaluated at the mean of all covariates in the case of the probit model, or simply the estimate $\hat{\theta}_M$ for the linear model. A similar metric has been used by French (2005). Cutler et al. (2013) calculate the decline in employment not explained by declining health.

Θ_M is a function of the estimated parameters. In the probit specification, it is the estimated marginal effect of health or cognition evaluated at the mean of all explanatory variables. In the linear model, Θ_M equals the estimate of θ_M, $\hat{\theta}_M$ where M denotes the corresponding measure of health or cognition.[ix]

IV   Data and descriptive statistics

This paper uses waves 1 to 6 of the English Longitudinal Study of Ageing (ELSA), covering years 2002–12, and waves 3 to 11 of the US Health and Retirement Study (HRS), covering years 1996–2012. We excluded the first two waves of HRS because of non-negligible changes in the questionnaire that happened in wave 3. Moreover, it is the later version of the HRS that informed the design of ELSA, so it is for these waves where the two surveys are most comparable. In both cases, the sampling is designed to become representative of the population aged 50 or older of their respective countries as the survey matures. Both HRS and ELSA collect biannual longitudinal data on respondents and their spouses, for the latter irrespective of their age, on a vast range of socio-economic, demographic, health and cognition variables.

ELSA respondents are a subsample of the Health Survey for England (HSE) in 1998, 1999 or 2001, representing the population of non-institutionalized individuals living in England and aged 50 or older in 2002/2003. Later interviews were conducted in 2004/2005, 2006/2007, 2008/2009, 2010–11 and 2012–13, with booster samples every 6 years.

The HRS began in 1992, with a representative sample of non-institutionalized individuals living in the United States aged 51 to 61 and their spouses. These individuals were interviewed biannually, even when later admitted to nursing homes (although, for consistency with ELSA, we exclude those in nursing homes), and refreshment samples were added every 6 years. We augment the HRS dataset with the RAND HRS Data File which contains cleaned versions (including some minor imputations) of the core HRS variables.

Throughout the paper, we focus on the retirement period using data for respondents and their spouses aged 50–70. Sample sizes for our population of interest are outlined in Table 1. Increases in waves 3 and 6 in ELSA and 4, 7 and 10 in HRS are due to refreshment samples. The overall sample size in the HRS is more than twice that for ELSA, due to both the larger

number of waves and the larger number of individuals in each wave. The total number of observations reported at the bottom row of Table 1 represents individual × time observations.

Our analysis separates three educational groups: College degree or equivalent, High School degree or equivalent (GCSE or A level in England), and High School Dropout (no GCSE qualifications in England).[x] We use the American labels in all future references. Figure 1 plots education levels against date of birth year for men aged 50 to 70 in ELSA and the HRS (Figure 2 shows the equivalent figures for women). The education composition of the English labor force changed considerably over these cohorts, with the proportion of men who at least graduated from High School increasing from about 35% among those born in the early 30s to about 80% among those born in the early 60s. English women departed from a lower basis of about 20% but reached similar education levels to those of men in the later cohorts.

Although the younger cohorts born in the 1960s look very similar across the two countries, there are important differences in the education achievement of older cohorts; education levels are much higher in the US than England for the older cohorts. In contrast, men and women from the younger cohorts are more likely to graduate from college in England than the US and are equally likely to leave school without qualifications. It is therefore important to bear in mind that individuals lacking any qualification in HRS are likely to be from lower in their country's skill distribution than their counterparts in ELSA.

The two surveys contain life history information that we use to describe permanent individual characteristics that drive both health, cognition and employment outcomes. Specifically, we use historical data on health during childhood and accumulated years of working experience in first observation to capture long-term health status and labor market attachment. These variables complete the set of initial conditions we control for, which also include health, employment, marital status and non-housing wealth observed when each individual first joins the sample.

A   Employment Profiles

We now turn to our key outcome variable, employment. Figure 3 shows significant declines in employment for all three education groups for both genders, particularly after age 60. In ELSA, employment among men starts from a higher base than that of women, and declines later; a sharp decline coincides with the State Pension Age (at 65 for men, 60 for women) in both groups. In contrast, both men and women experience similar declines in employment rates with age in the US, where the Early (62) and Normal (66 for most of the sample period) Retirement Age is the same for the two genders. These profiles for the two countries are suggestive of the importance of retirement incentives in driving the decline in employment. Employment rates are flatter in the HRS than in ELSA, implying that a higher proportion of Americans than English are still working in their late 60s. Finally, the education gradient is much stronger in the US than it is in England. Fewer High School Dropouts are in work during their 50s in the US than England. This feature is likely to be linked to the differences in education attainment of Americans and English, with High School Dropouts being a much larger, and hence probably less disadvantaged, group in England.[xi]

B   Objective measures of health

As described in the methods Section III, we consider health variables in two broad categories, objective and subjective. Here we focus on the former. Table 2 summarizes the objective health measures we consider, which include reports of the health conditions for which respondents receive medical treatment (such as cancer or diabetes). For comparability, we only use variables that are present both surveys.

The differences between the US and England are stark; prevalence in the US is larger for 8 out the 10 conditions for which the respondent is treated (top ten rows in the Table), and is often twice or even three times larger in magnitude. For example, cancer prevalence is 3% in ELSA for both men and women, but the figures in the HRS are, respectively, 8% and 11%;

diabetes prevalence is 9% and 6% for men and women in ELSA and is 19% and 17% in HRS; the numbers for arthritis are 23% and 34% in ELSA and up to 44% and 57% in HRS.

These reported health differences have been well documented before in Banks et al. (2006) and Banks et al. (2016). They may reflect a combination of differences across the two countries, in health status, diagnosing rates and respondents' information about their health conditions. Meanwhile, gender differences are similar across the two countries; typically, women are more likely to have arthritis and psychiatric problems, but are less likely to have suffered from a stroke, heart attack or diabetes.

Panels A and C of Figure 4 show how the prevalence of arthritis changes between the ages of 50 and 70, by gender and education in England and the US. The plotted lines show smoothed age trends using a moving averages of 3 years. The clear positive gradient with age for all groups is indicative of how health deteriorates around the retirement age. This unsurprising finding justifies the focus on this age group of much of the economic literature on health and employment in developed countries. The graphs also show that the prevalence of arthritis is higher among women and those with less education in both countries. The latter is also typical of many health conditions: less educated and poorer individuals tend to report lower levels of health. However, the sharpest difference is that between England and the US, with arthritis being much more prevalent for all groups in the US.

These figures may mask cohort differences in the prevalence of the disease. To deal with this, we net out fixed effects by estimating

$$h_{it} = \alpha_i + \beta_t + u_{it}$$

where h_{it} is a health outcome of interest for individual i aged t, α are the individual fixed effects (normalized to have mean zero in the population), and β_t are a full set of age dummy variables that capture health-age profiles net of fixed effects. We then plot the estimated age profile β_t. Note that this fixed effects specification captures all time invariant factors. For example, a

cohort effect is just the average fixed effect of everyone within that cohort. In our application it is important to net out fixed effects particularly when looking at health profiles conditional on education because of the rapid increase in education attainment over the sample period, especially in England. Specifically, the shift towards more education implies that highly educated individuals in the older cohorts of our sample may be drawn from a more selected sample, with different health outcomes, than equally educated individuals from the younger cohort. The fixed effects estimator, which is identified by individual changes in health with age, eliminates the effects of such compositional changes on the level of health. In addition, because fixed effects tracks the same people over time, it addresses the issue of non-random attrition from the sample due to death or other reasons. Profiles for arthritis are shown in Panels B and D of Figure 4, respectively for England and the US. The patterns are similar to those in the raw data, but the age gradient is noticeably steeper for most groups. The full set of figures describing the prevalence of health outcomes by age is available in Section 2.2 of the Online Appendix.

C Subjective measures of health

The indicators of subjective health are summarized in Table 3. These are variables of self-reported health, describing general health and whether it hinders work or the ability to perform normal daily activities. The means reported in the table show some interesting patterns. Responses to all questions are well aligned across the two countries, with English people reporting slightly better health than Americans but with much more modest differences than those observed for objective health measures. This is remarkable given the considerably higher prevalence of disease in the US as described by the objective measures. It must be driven, at least to an extent, by large differences between the two countries in the way individuals report their own health. This is consistent with results in Banks et al. (2016) showing that Americans set lower thresholds for good and excellent self-reported health than do the English, and in

Kapteyn et al. (2007) showing that Americans set lower thresholds for being non-disabled than the Dutch.

Finally, the English tend to report lower levels of health as children than Americans do, with around 12% of ELSA respondents reporting bad health as child compared to 7% of HRS respondents.

We summarize the subjective measures of health in a single index that we think captures well the global measure of health status, the first component from a Principal Component Analysis of the three subjective health measures.[xii] The age profiles of the index are shown in Figure 5. The patterns are much more similar across the two countries than those found for the objective measures. There is again a clear ordering by education group and a negative gradient with age. Removing fixed effects changes the patterns for the US more than it does for England, by making the age profiles steeper.

D   Cognition

High quality survey information on cognitive functioning only recently started to become available. It exists in both ELSA and HRS, with respondents being given a battery of cognitive tests. The literature on cognitive skills in adults (e.g. Choi et al., 2014) has distinguished between measures of crystallized intelligence (which relies on accessing information from long-term memory) and fluid intelligence (the capacity to think logically and solve problems in novel situations, independent of acquired knowledge).[xiii] Our focus is on fluid measures, primarily because they are available in both surveys across several waves,[xiv] though also because previous studies have found that it is fluid and not crystallized intelligence that is positively correlated to labor outcomes (for example, Anger and Heineck, 2010, and Heineck and Anger, 2010).

Both datasets include several cognitive measures of fluid intelligence. We focus on two of the tests in the survey alongside two of the Instrumental Activities of Daily Living (IADL)

measures which also reflect cognition. The measures are summarized in Table 4. The table shows that Americans do slightly worse in cognition tests than the English, with 10% (respectively 3%) reporting difficulty using a map, 4% (2%) reporting difficulty managing money, and average scores of 5.8 (6.1) and 4.8 (4.9) out of 10 in the recall and delayed recall tests.

Similar to the construction of our health index, we construct a cognition index that summarizes the information content of the four cognition variables using Principal Component Analysis. The first principal component is plotted in Figure 6.[xv] In general, there is a clear worsening in cognition with age as assessed by this test. What is remarkable, however, is that the age profiles in ELSA are essentially flat once fixed effects have been removed (Panel B). This suggests that the deterioration in cognitive skills with age seems to be explained by compositional changes across cohorts in England: older individuals have lower cognition not because of their age, but because they were born into older cohorts with lower cognition over their life.[xvi] The figure also shows evidence of a clear ordering by education group in the scoring of the recall tests, with the highest educated scoring best and the lowest educated scoring worst. Moreover, the gap between the high educated and the low educated is considerably larger in the US.

5   Empirical results

In this section we compare the estimates of the impact of health on employment using various specifications commonly adopted in the literature. We use subjective health measures, either on their own or combined in an index, and we extend the model to include cognition. We show the importance of allowing for initial conditions when estimating the impact of health. We address the issue of measurement error in health using instrumental variables, and demonstrate that the linear regression model predicts accurately the impact of health on employment. And finally, we explore the differences in results between England and the US. For conciseness, we

focus on estimates based on the latent index probit model in equation (2) and show the very similar findings we obtained for the linear probability model in Section 4.4 of the Online Appendix. The effects of health and cognition on employment are calculated using the marginal effects at the average of all regressors included in each model.

A   The Effect of Subjective Measures of Health and Cognition on Labor Supply

Table 5 displays estimates of the effects of a one standard deviation improvement in the health or cognition indexes on employment. As described in the previous section, the subjective health index is the first principal component of the three subjective health measures and the cognition index is the first principal component of the four cognition measures. Each cell in Panels A and B reports estimates from a separate regression; cells in the top and bottom halves of Panel C report, respectively, the cognition and health coefficients in regressions that control for both. Sample sizes are shown in the bottom panel.

The relationship between subjective health and employment is shown in Panel A. Estimates in Column 1 are for men in England; they are obtained from a set of education-specific regressions of employment on the subjective health index and a basic set of controls that only includes a quadratic polynomial in age and year dummies. In ELSA, a one standard deviation improvement in the subjective health index is associated with 17.7% higher employment amongst high school dropout men; comparable estimates for high school graduates and college graduates are 11.0% and 7.1%, respectively.

However, estimates of the effects of subjective health on employment may be biased by unobserved factors that relate to both. For instance, individuals from poor backgrounds may have missed on the critical investments that foster good health as well as other skills required in work environments. If poor health and unobserved skill deficits lower employment rates later in life, then failure to control for skill will confound estimates of the employment effects of health. To deal with this sort of problem, we add a full set of initial conditions to the

31

regression model, including health status during childhood, accumulated years of working experience, as well as health, cognition, employment, marital status and non-housing wealth when first observed in the sample. These variables capture existing heterogeneity at the start of the observation period that relates to both employment and health.

For men in ELSA, the new set of estimates controlling for initial conditions can be found in Column 2. The reported coefficients in Panel A measure the impact of changes in health on changes in employment during later working years. The effects of health roughly halve with the inclusion of initial conditions in the regression model, showing that indeed much of the relationship between health and employment among English men is spurious. We find very similar patterns for English women (see Panel A, Columns 5 and 6), although with estimates that are generally slightly smaller. HRS estimates, meanwhile, are modestly larger than ELSA estimates but are less affected by the inclusion of initial conditions (Columns 3–4 and 7–8 for men and women, respectively).

Panel B shows equivalent estimates for the effects of cognition. These are always smaller than the effects of subjective health. In ELSA, a one standard deviation improvement in the cognition index of men is associated with 8.7%, 3.3% and 1.3% higher employment rates among high school dropouts, high school graduates and college graduates, respectively (Column 1, Panel B). Adding initial conditions to the regression model, which now include the cognition index but not the health index in the first observation period, considerably reduces the estimated effects. HRS estimates are larger, and are again less affected by the inclusion of initial conditions. Estimates for women are very similar to those for men.

Panel C in Table 5 shows results for employment regressions on both the cognition and subjective health indexes. It shows that health remains a strong determinant of employment among older workers even when accounting for cognition, but that cognition plays a much more modest role (if any) after accounting for health. In line with findings in Panels A and B,

Panel C also highlights the importance of controlling for permanent heterogeneity when estimating the impacts of cognition and subjective health on employment. We therefore focus exclusively on estimates from regression models that include initial conditions in what follows.

Table 6 displays estimates of the share in employment decline between ages 50 and 70 that can be explained by a decline in health and/or cognition over the same period. It uses the coefficients in Table 5 to calculate the percentage change in employment explained (δ in Equation 13). Estimates in Column 1 of Panel A show that the deterioration in health explains between 4.0% and 7.2% of the decline in men's employment in ELSA. The impact is largest for the high school dropouts and falls with education. Column 1 in Panel C shows that these estimates are barely affected by the inclusion of cognition, in line with cognition having a negligible impact on the employment of older workers in England (see also Panel B). Contrasting Columns 1 and 3 in the Table shows that changes in health and cognition explain generally less of the changes in employment of women than men, particularly among those who leave education without qualifications.

Results for the HRS display similar patterns to those found in ELSA, only stronger (Columns 2 and 4 in the Table). In particular, they suggest that both health and cognition play a role in explaining the decline in employment of American workers near retirement age, though the impact of health decline is about two to four times larger than that of cognition decline (Panels A and B). Moreover, cognition explains about two additional percentage points of the decline in employment when added to health in the same regression model (Panel C versus A).

The incremental value of cognition is tested in Table 7. Figures in Columns 1 to 4 show the change in explained share of employment decline induced by adding cognition in addition to health, in percentage terms relative to the effect of health alone; these numbers are obtained from comparing estimates in Panel C and A of Table 6. Columns 5 to 8 show the *p*-values for

testing the equality between the same two sets of estimates, with and without cognition. The results suggest that cognition increases modestly the explained employment decline in the HRS but the differences are never statistically significant at a 5% level. In line with our earlier findings for ELSA, cognition plays no discernible role in driving employment in England.

By summarising the information on subjective health in a single index, we may be discarding important information. Our subjective health index is constructed using three variables. In principle, each of the three variables could have independent explanatory power for employment beyond their contribution to the index. To test whether this is the case, we estimated alternative empirical specifications of the employment regression model and used them to predict the share of employment decline driven by health over the same 50–70 age period (δ in Equation 13). Estimates are displayed in Table 8. Panel A reproduces Panel A in Table 6 and is the reference set of estimates, obtained using the single subjective health index. Panel B adds all three measures of subjective health separately to the employment regression; this has little effect on the estimates.[xvii] Panel C includes only one of the subjective health variables directly measured in the questionnaire, the dichotomous variable for whether health limits work; estimates of the δ's are modestly lower in this case, suggesting that this single measure misses some of the drivers of employment, or that there is significant measurement error in the variable. Section 4.3 of the Online Appendix shows estimates using the other subjective measures individually. The individual subjective measures always produce smaller and more variable estimates of the impact of health than the health index using all three measures. This suggests that a single health index, if properly constructed, is sufficient for capturing the effect of health on employment; however, a single subjective measure is not sufficient.[xviii]

Table 9 further quantifies the importance of accounting for more detailed subjective health information by comparing Panels B and C with Panel A of Table 8. Columns 1–4 detail the

percentage differences between the estimates in these panels, using estimates in Panel A as baseline, while Columns 5–8 detail the *p*-values for testing their equality. The figures in the top panel reveal that the relative differences induced by fully accounting for the subjective health information are generally small and mostly negative. In most cases we fail to reject equality; in some cases we do reject, but the only rejection of a positive difference (which would indicate that the three measures separately contain more information for employment than the composite index) is for women with high school diploma in the HRS, for whom the relative difference is very modest.

However, the inspection of the bottom panel in Table 9 reveals that the information in a single observed measure significantly under-represents the variation in subjective health relevant for employment, particularly in ELSA. For all groups in ELSA, the share of employment decline explained by changes in this measure is at least 50% lower than the same measure for the subjective health index. For the HRS, the use of the single measure 'health limits work' also produces smaller effects of changes in health on employment than those produced by our health index, but the differences are smaller and only statistically significant at conventional levels for men.

Overall we find that the single subjective health index captures the variation in health that is responsible for the decline in the employment rates of older workers as well as more detailed measures of subjective health do. Our parsimonious yet complete representation of health is particularly useful in contexts that are only practical with low-dimensional specifications, such as in structural models of health, employment and earnings. We therefore focus on results based on the single subjective health index in what follows.

B Using Instrumental Variables to Address Justification Bias and Measurement Error in Subjective Health Measures

Subjective health measures can be afflicted by justification bias and measurement error that confound estimates of the effects of health on employment if subjective health is used as a proxy for health status. We address this problem by instrumenting it with the full set of objective measures. Objective measures focus on specific conditions and thus may provide an incomplete picture of health status, but they are likely to be strongly related to the subjective measures. Moreover, measurement error and justification bias in subjective health is likely to be unrelated to objective health. These features make the objective measures an ideal candidate for instrumenting the subjective health index. Since the direction of the bias resulting from using subjective health to proxy health status is indeterminate a priori, so is the direction of the correction from instrumenting it (see discussion in Sections III.B and III.C): IV estimates should be smaller than their linear counterparts if justification bias dominates, while the opposite holds if attenuation bias dominates.

We start by testing the strength of the instruments when using the entire set of objective measures, and will then discuss how estimates of the effects of health on employment change with instrumenting. To test for weak instruments, we compare the F-statistics to Stock-Yogo critical values: we reject the null of no statistically significant relationship between the subjective health index and the objective health measures at the 5% significance level for all gender $\times$ education $\times$ country cells, whether or not cognition is included in the regression model of employment. This demonstrates that the objective measures are strong predictors of the subjective health index.

IV estimates of the fraction of employment decline explained by health and cognition are shown in the two panels of Table 10, Panel A for the impact of health only and Panel B for the joint impact of health and cognition. The estimates in both panels are very close; they are also

overall similar to the OLS estimates of the impact of subjective health and cognition on employment in Table 8. They reveal that declining health can explain at most 15% of the decline in employment around retirement age, and that cognition adds little to this and only for the HRS. What is also apparent from these estimates is that both health and cognition are stronger drivers of the employment choices for Americans than for the English. We further discuss this point in Section V.D.

The two panels of Table 11 compare the IV estimates in Panels A and B of Table 10 with their OLS counterparts, respectively in Panels A and C of Table 6; the first four columns show the relative differences between the IV and OLS estimates, using OLS estimates as the baseline, and Columns 5–8 show the p-values for testing their equality. The results suggest that measurement error and justification bias do not seriously affect estimates, or at least that they offset. The OLS estimates are of similar order of magnitude, albeit systematically smaller (hence the positive differences in Columns 1 to 4), than similar IV estimates. The null hypothesis that the OLS and IV estimates are equal is not rejected at conventional levels in most cases. In the couple of cases where it is rejected, which are both in the HRS, IV estimates are larger than their OLS counterparts.

For the IV approach to be valid, any measurement error affecting the objective measures must be orthogonal to that affecting the subjective measures. In particular, this rules out justification bias affecting both objective and subjective measures. It also rules out the possibility that detection of objective health conditions may be related to economic conditions, which might be the case if seeking medical attention is a choice affected by access to health insurance, or if those with higher socioeconomic status are more likely to be aware of their health problems (e.g. Johnston et al., 2009), for example. We test the validity of the IV approach by restricting the number of objective instruments to represent only major conditions. These major conditions usually require medical attention, making it unlikely that people would

wrongly report whether they suffer from one of them. The results from these estimates are shown in Section 4.3 of the Online Appendix and are not statistically different from the estimates using the full set of instruments. We therefore conclude that the measurement errors in our objective health measures and subjective health index are unlikely to be correlated. Our findings suggest that justification bias, which has been a major concern in the literature and is expected to bias estimates of the impact of health upwards, is either not very important or is more than compensated by attenuation bias from measurement error in the subjective measures.

Table 12 provides additional evidence on the validity of the single index assumption using the overidentification restrictions supplied by the many instruments we are using. If the objective measures affect employment only through their effect on subjective health, then the IV residuals should not be systematically related to any of the objective health measures.

We implemented the test by regressing the IV residuals on all the objective health measures and all other explanatory variables in the employment regression, and then calculating the F-Statistic for the full set of objective measures (as suggested by Davidson and MacKinnon, 2003; see equations 9 and 10).[xix] The residuals were clustered at the individual level to account for serial correlation. In Table 12 we show the *p*-values for testing the null hypothesis that objective measures affect labor supply only through the subjective health (the IV exclusion restriction). The test results show that the exclusion restriction is rejected in the majority of the cases in the HRS, whether or not cognition is included in the regression model, but it is never rejected with the ELSA data.

One possibility is that the impact of health on employment varies with health conditions, in line with the argument that it is the serious and persistent conditions that most affect employment. We test whether this may be the case by restricting the objective instruments to a subset of major health conditions. These are heart problems, lung disease and whether the individual has suffered a stroke or heart attack. When re-running the test on these more

homogeneous set of conditions we find much stronger support for the single index assumption. Table A17 in the Online Appendix shows that, whether or not cognition is included in the regression, the null is only rejected for three out of twelve cases (in all cases, better educated individuals from the HRS). This result suggests that the impact of changes in health may be more important if these are driven by the onset of more serious (and potentially long-lasting) health conditions.

C   Assessing bias due to omitted objective health measures

Objective health information is only collected for a subset of the relevant conditions, which is likely to result in downward biased estimates of the impact of health on employment as discussed in Section III.A. Here we assess the bias when using only a limited set of objective measures to proxy for health. We estimated the alternative model of health as a function of the entire set of objective measures in equation (4) to assess the severity of bias due to omitted objective measures; estimates using all objective measures can be found in Panel D of Table 13. Even when they are added in a fully flexible format, all objective measures together predict an employment decline that is generally smaller than the estimated effects based on the subjective heath index – see Table 14 for percent differences and p-values for testing the equality of predicted share in employment decline explained by objective and subjective measures. The differences are modest, although statistically significant for many groups, particularly in the HRS. For high school dropout women in both ELSA and the HRS, the share of the employment fall predicted by health is actually larger when using the full set of objective measures than when using the subjective health index; however, the differences are small and only statistically significant for the HRS data.

These results are consistent with the hypothesis that using only a limited set of objective measures provides an incomplete view of the health status affecting work capacity, and underestimates the impact of health on employment (recall discussion in Section III.A). More

generally, however, our predictions of the effects of health based on objective and subjective measures are much more similar than has been suggested in previous studies. Existing estimates based on objective measures used only a subset of the measures we use here and found that they produced much smaller estimates than subjective IV estimates: Bound (1991), for example, found that, a single objective measure (future mortality) produced estimates of the effect of health that were only about one tenth of the size of the subjective or IV estimates. Interestingly, but perhaps predictably, we now find that a comprehensive set of objective health measures available in the HRS and ELSA produces estimates that are much closer to the subjective IV estimates.

To further investigate the effects of using limited subsets of objective health measures, Panels A to C of Table 13 show estimates of the explained share of employment decline from regressions that gradually add more objective measures. The set of estimates in Panel A are based on a single health measure, specifically whether the individual reports that they have high blood pressure; estimates of the impact of health on employment in this specification are very small and are not statistically significant at conventional levels. These results align well with the findings in Bound (1991).[xx] The surprising results, however, are in Panel B. They show that the estimates of the impact of health quickly converge to levels very close to those obtained when using the full set of objective measures by adding just three more measures of objective health that arguably capture a wide range of conditions (arthritis, psychiatric and lung diseases). Further adding more conditions does not much change the estimates (Panels C and D).

D   Exploring between-country differences

Our estimates show that the share of decline in employment that is explained by declines in health is consistently greater in the US than it is in England for all groups, often larger by a factor of approximately three. Here we decompose the differences in our main set of estimates

for the US and England – the δ parameters defined in Equation (13) and presented in Table 10. Table 15 uses an Oaxaca decomposition to describe how much of the difference $\delta_{US} - \delta_{ENGLAND}$ is explained by differences in the impact of health and cognition on employment (θ), differences in deterioration in health and cognition (ΔH) and differences in the employment decline (ΔY). Breakdowns are provided for both sets of estimates from Table 10, depending on whether only health (Panel A) or also cognition (Panel B) are accounted for in estimating δ.[xxi]

The general picture for all cases is that the majority of the between-country differences in how much of the decline in health is explained by health or health and cognition can be attributed to differences in the impact of these variables on employment (θ); differences in the decline of health, cognition and employment are less relevant. The role of the impact of health on employment is particularly dominant among men with less than college education, for whom it drives almost the entirety of the between countries difference. For other groups, across countries differences in θ's explain two thirds or more of the differences in δ's.

The larger response of employment to health in the US may result from differences in the institutional backgrounds of the two countries shaping the employment responses to health around retirement age. For instance, the two countries differ in the provision of health insurance, which is universal in England but not in the US, the generosity of disability benefits and the rigor of its entitlement rules, and the design of financial incentives to retire and their age-dependence. For example, the US disability system, which provides a health dependent benefit, is more generous than the English one, and provides benefits only if beneficiaries do not work. Thus unhealthy Americans have a strong incentive not to work. Compared to the US, England provides more generous out of work benefits for reasons unrelated to health, such as unemployment benefits. All these institutions are expected to play an important role in determining retirement choices and their dependence on health. While establishing the

importance of these channels certainly merits further research, this is beyond the scope of this paper.

Less than one quarter of the difference for men, but more than one quarter of the difference for women, can be explained by a larger employment drop in England among those in their 50s and 60s. Here we notice that employment drops sharply in England at the state pension age (60 for women, 65 for men), but it declines much more gradually and slowly in the US (recall Figure 3). While this is likely related to differences in the retirement incentives for these age groups, it implies that Americans are more likely to work into older ages than the English. Hence, Americans may be more exposed to the onset of health conditions leading to retirement during their (longer) working lives. In turn, the English are more likely to be already retired when experiencing a similar deterioration in health.

VI   A framework to understand the employment choices of older workers

The previous section presents reduced form evidence that bad health is associated with lower employment, conditional on past employment, health, and other variables. This section presents an economic model of employment choices, savings and health for older workers, and use it to motivate the empirical strategy used in the previous section, to highlight its underlying assumptions, to guide the interpretation of our estimates and to discuss the key mechanisms that drive the impact of health on employment. The structural model can be used to identify and quantify such mechanisms (e.g. preferences, productivity, and financial incentives).

A   The model

We consider the problem of individuals deciding whether to work near retirement age. Our aim here is to focus on the simplest dynamic model that can represent the many ways in which health affects employment among older workers. In our model, individuals decide in each period whether to work, how much to consume and how much to save for the future. They do

so in a risky environment, where they face uncertainty in future health, wages and preferences for working. Health-related benefits, or disability benefits, partially insure against income losses associated with bad health, but as with other social insurance instruments, they also change working incentives. Individuals may save to further insure themselves against economic consequences of health and other shocks. In what follows, we briefly formalise the model.

Preferences

Individuals are indexed by *i*. They seek to maximize the expected discounted value of their present and future utility by choosing employment at each age *t*. We consider a single cohort, so age and time are used interchangeably. In each period, workers derive utility from consumption and leisure in a way that depends on health status:

(14) $$u(C_{it}, Y_{it}, H_{it}, \xi_{it}, \zeta_{it}) = \left(Y_{it} * (\theta_0 + \theta_1 H_{it} + \xi_i + \zeta_{it})\right) + \frac{C_{it}^{1-\gamma}}{1-\gamma}$$

where *u* is the per-period utility function, *C* is consumption, *Y* is employment status and assumes the values 0 and 1 for not-working and working respectively, and *H* is health. Including health in the utility function formalizes the idea that working is more costly in periods of poor health and captures the empirical regularity that sick people work less. Finally, ξ and ζ represent unobserved idiosyncratic permanent and transitory preferences for work, respectively.

Budget Sets

The potential earned income of individual *i* at age *t*, W_{it}, is realized if $Y_{it} = 1$. It varies with age and health status. For simplicity, we omit other exogenous characteristics that may drive wages. We allow for two individual-level unobserved components in wages, a permanent unobserved heterogeneity element ϕ, which we interpret as ability, and a time-varying wage shock v. Earned income is, therefore:

(15)
$$W_{it} = \omega(t, H_{it}, \phi_i, \nu_{it})$$

Individuals in bad health may be eligible to benefits $B_t(H_{it})$ but entitlement depends on their other income, being taxed away at a rate $\tau_t(W_{it}, Y_{it}, H_{it})$, which may change over time and with the age of the individual as s/he approaches retirement. So the asset accumulation equation is:

(16)
$$A_{it+1} = (1+r)[A_{it} + W_{it}Y_{it}(1 - \tau_t(W_{it}, Y_{it}, H_{it})) + B_t(H_{it}) - C_{it}]$$

Health and Mortality

When deciding about employment and savings, individuals are faced with health uncertainty. We pose that health follows an age-dependent Markov process

(17)
$$H_{it} = h(t, H_{it-1}, \psi_i, \epsilon_{it})$$

where ψ and ζ are the unobserved permanent and transitory elements of health. In line with our empirical findings, we model health as a uni-dimensional variable.

Besides its impact on the utility cost of work and wages, we also formalise the impact of health on survival: a worker alive at age t with health status H survives to age t+1 with probability $s(t, H)$.

Structure of the unobserved components

We allow for unobserved heterogeneity in health, wages and preferences for work (ψ_i, ϕ_i, ξ_i) and for arbitrary correlation between these three dimensions of heterogeneity. We also consider transitory unexpected shocks to health, wages and preferences, (ϵ_{it}, ν_{it}, ζ_{it}) which are serially uncorrelated, mutually independent and independent from the unobserved heterogeneity components.

The individual's problem

At age t, the state vector of the worker i is $\Omega_{it} = (H_{it}, W_{it}, A_{it}, t, \psi_i, \phi_i, \xi_i, \zeta_{it})$. In recursive form, the worker's problem is

(18)
$$V_t(\Omega_{it}) = \max_{C_{it}, Y_{it}} \{ u(C_{it}, Y_{it}, H_{it}, \xi_i, \zeta_{it}) + \beta s(t, H_{it}) E_t V_{t+1}(\Omega_{it+1})$$

subject to equations (15) – (17). In the above equation β, is the subjective discount factor.

B   How health affects employment of older workers

The addition of health to an otherwise stylized structural model of employment and savings exposes various channels through which health affects employment. In our simple model, a negative health shock reduces preferences for work, wages and expected longevity, and it increases entitlement to benefits.[xxii]

Formally, the structural labor supply function is

$$(19) \qquad Y_{it} \ = \ Y(H_{it}, W_{it}, A_{it}, t, \psi_i, \phi_i, \xi_i, \zeta_{it} | \theta)$$

where θ is the set of all parameters in equations (14) – (18). In the context of this labor supply function, we can see multiple pathways by which health deterioration with age may impact labor supply. In particular, we identify five channels through which bad health shocks can discourage work, all of which are expressed in our structural model:[xxiii]

1. Preferences. Bad health can raise the marginal utility of leisure relative to that of consumption (Capatina, 2015). This is embodied in equation (14), where health is allowed to interact with the utility value of working. For this reason, health *H* impacts employment directly in equation (19).

2. Productivity. Bad health can lower workers' productivity and resulting wages. This is represented in equation (15), where function ω(.) captures the potentially negative impact of health on wages. So health may affect employment indirectly through wages *W*.

3. Disability insurance benefits. People in sufficiently bad health may qualify for benefits from disability programs. This is embodied in *B*, the benefit amount, and also in *τ,* the share of earned income that is taxed away. Those receiving benefits have incentives to reduce labor supply through several channels. First, the benefits provide income,

allowing individuals to purchase more leisure. Second, in many countries the benefits are means-tested and sometimes limit work altogether. Moreover, in the US, beneficiaries can receive Medicare or Medicaid health insurance depending on their working income, with excessively high income triggering the loss of benefits.

4. Expectations of future employment and earnings capacity. The persistent health process described equation (17) implies that current shocks may have long-lasting effects on future health and thus future employment and earnings capacity. This changes the value of savings and, hence, that of employment.

5. Life expectancy. With shorter expected lifespans, individuals in bad health may not need to work as long to accumulate savings for retirement. This effect operates through the survival probability *s(t, H)* in equation (18).

Most papers consider only a subset of these channels. For example, French (2005) and Capatina (2015) consider four of the five channels, excluding only disability benefits. French et al. (2018) and Kitao (2014) accounts for disability benefits but French et al. (2018) use a stylized model of disability benefits and Kitao (2014) uses a very stylized model of demographic transitions and health insurance.

C   Approximation Model

In this section we demonstrate how our simple reduced form model of employment, described in detail in Section III, can be derived as an approximation to the solution of the dynamic labour supply model. The process of doing so provides further clarity on the interpretation of the estimates in Section V.

In the structural model, the work decision is defined as

$$(20) \quad \mathbf{1}\{Y_{it} = 1 | \Omega_{it}\} = \mathbf{1}\left\{\max_{C_{it}}\{u(C_{it}, 1, H_{it}, \xi_i, \zeta_{it}) + \beta s(t, H_{it})E_t V_{t+1}(\Omega_{it+1})\} - \right.$$

$$\left.\max_{C_{it}}\{u(C_{it}, 0, H_{it}, \xi_i, \zeta_{it}) + \beta s(t, H_{it})E_t V_{t+1}(\Omega_{it+1})\} > 0\right\}$$

To the extent that time periods are short, separability between leisure and consumption in the utility function implies that the marginal utility of consumption is only mildly affected by labour supply. This simply reflects consumption smoothing, as any additional income received in a period is consumed over time and, therefore, mostly saved in the period it is realized. But then, the additional income from work will be valued at the marginal value of assets (which, by the envelope condition, equals the marginal utility of consumption). We can then rewrite equation (20) as

$$\mathbf{1}\{Y_{it} = 1|\Omega_{it}\}$$

$$\approx \mathbf{1}\{(\theta_0 + \theta_1 H_{it} + \xi_i + \zeta_{it}) + \beta(1+r)s(t, H_{it})W_{it}(1 - \tau_{it}) \frac{\partial E_t V_{t+1}(\Omega_{it+1})}{\partial A_{it+1}}$$

$$= \mathbf{1}\{(\theta_0 + \theta_1 H_{it}) + C_{it}^{-\gamma} W_{it}(1 - \tau_{it}) + \xi_i + \zeta_{it} > 0\}$$

$$= \mathbf{1}\{Y_{it}^* > 0\}$$

where $C_{it}^{-\gamma}$ is the marginal utility of consumption, $W_{it}(1-\tau_{it})$ is the change in income induced by a move into work, τ_{it} is an abbreviation for $\tau_t(W_{it}, Y_{it}, H_{it})$ and Y_{it}^* is the latent employment index. This is the discrete choice version of the marginal rate of substitution condition, a condition that holds exactly as the time periods become arbitrarily short.

In a cross section, H_{it} may be correlated with $Y_{it}, W_{it}, \tau_{it}$, or ξ_i, leading to biased estimates of θ_1 if these variables are not added to the regression model. This is for two reasons. First, while θ_1 is a deep parameter that represents how preferences for work change with health, its estimate will conflate other mechanisms such as the indirect impact of health on employment through its effect on wages. Second, health is likely correlated with permanent individual characteristics that also determine employment, such as those settled in childhood through investments and other factors. We use initial conditions to address this second problem.

To proceed, we write the employment index for period t and an initial period 0:

(21) $$Y_{it}^* = \theta_0 + \theta_1 H_{it} + C_{it}^{-\gamma} W_{it}(1 - \tau_{it}) + \xi_i + \zeta_{it}$$

$$(22) \qquad Y_{io}^* = \theta_0 + \theta_1 H_{i0} + C_{i0}^{-\gamma} W_{i0}(1 - \tau_{i0}) + \xi_i + \zeta_{i0}$$

We then combine these two equations to obtain the following expression:

$$Y_{it}^* = \theta_1 (H_{it} - H_{i0}) + \left(C_{it}^{-\gamma} W_{it}(1 - \tau_{it}) - C_{i0}^{-\gamma} W_{i0}(1 - \tau_{i0}) \right) + (\zeta_{it} - \zeta_{i0}) + Y_{i0}^*$$

$$(23) = \theta_1 (H_{it} - H_{i0}) + \left[\left(\frac{C_{it}}{C_{i0}} \right)^{-\gamma} \frac{W_{it}(1 - \tau_{it})}{W_{i0}(1 - \tau_{i0})} - 1 \right] C_{i0}^{-\gamma} W_{i0}(1 - \tau_{i0}) + (\zeta_{it} - \zeta_{i0}) + Y_{i0}^*$$

$$\approx \theta_1 (H_{it} - H_{i0}) + \left[-\gamma \ln \left(\frac{C_{it}}{C_{i0}} \right)^{-\gamma} + \ln \left(\frac{W_{it}}{W_{i0}} \right) + \ln \left(\frac{(1 - \tau_{it})}{(1 - \tau_{i0})} \right) \right] C_{i0}^{-\gamma} W_{i0}(1 - \tau_{i0}) + (\zeta_{it} - \zeta_{i0}) + Y_{i0}^*$$

where the third approximate equality results from a simple Taylor series approximation. The key issue to notice here is that, by using initial employment we were able to eliminate unobserved heterogeneity in preferences for work from the employment equation.

We now project growth rates in the marginal utility of consumption, wages, and taxes, weighted by the initial marginal of consumption and the initial after tax wage, on initial health, change in health, initial employment index and other exogenous variables:

$$-\gamma \ln \left(\frac{C_{it}}{C_{i0}} \right)^{-\gamma} C_{i0}^{-\gamma} W_{i0}(1 - \tau_{i0}) = \delta_{C0} + \delta_{CH}(H_{it} - H_{i0}) + \delta_{CH0} H_{i0} + \delta_{CY0} Y_{i0}^* + \delta_{CZ} Z_{it} + \omega_{Cit}$$

$$\ln \left(\frac{W_{it}}{W_{i0}} \right) C_{i0}^{-\gamma} W_{i0}(1 - \tau_{i0}) = \delta_{W0} + \delta_{WH}(H_{it} - H_{i0}) + \delta_{WH0} H_{i0} + \delta_{WY0} Y_{i0}^* + \delta_{WZ} Z_{it} + \omega_{Wit}$$

$$\ln \left(\frac{1 - \tau_{it}}{1 - \tau_{i0}} \right) C_{i0}^{-\gamma} W_{i0}(1 - \tau_{i0}) = \delta_{\tau 0} + \delta_{\tau H}(H_{it} - H_{i0}) + \delta_{\tau H0} H_{i0} + \delta_{\tau Y0} Y_{i0}^* + \delta_{\tau Z} Z_{it} + \omega_{\tau it}$$

In the above expression, Z_{it} summarises variables that may affect changes in consumption, wages or taxes, including time dummies and a second order age polynomial. Thus the coefficients on age in the projection of the rate of change in the tax rate (line 3) will partly capture how the work incentives change with age around retirement.

By replacing the above expressions in equation (23) yields the key equation that motivates our empirical specification:

$$(24) \qquad Y_{it}^* = \delta_0 + \theta_H (H_{it} - H_{i0}) + \delta_{H0} H_{i0} + \delta_{Y0} Y_{i0}^* + \delta_Z Z_{it} + \omega_{it}$$

where

$$\delta_0 = \delta_{C0} + \delta_{W0} + \delta_{\tau 0}$$

$$\theta_H = \theta_1 + \delta_{WH} + +\delta_{CH} + \delta_{\tau H}$$

(25)
$$\delta_{H0} = \delta_{WH0} + +\delta_{CH0} + \delta_{\tau H0}$$

$$\delta_{Y0} = 1 + \delta_{CY0} + +\delta_{WY0} + \delta_{\tau Y0}$$

$$\delta_Z = \delta_{CZ} + +\delta_{WZ} + \delta_{\tau Z}$$

$$\omega_{it} = (\zeta_{it} - \zeta_{i0}) + \omega_{Cit} + \omega_{Wit} + \omega_{\tau it}$$

The second line of equation (25) shows that θ_H in equation (24) measures a combined effect of the change in health on employment, arising both from its impact on preferences to work and, indirectly, through its impact on the marginal utility of consumption, wages and the tax rate. Here the mechanisms discussed in the previous section are all represented: the direct impact through preferences is captured by θ_H; indirect effects through productivity and benefit entitlement are reflected on the parameters from the wages and tax projections, respectively; and changes in expected future health, lifespan and consequent future value of work are reflected in the parameters from the projection of the marginal utility of consumption.

The final line of equation (25) shows that the residual in equation (24) is a function of the transitory shocks to preferences for work, and the orthogonal residuals from the projection of the changes in the marginal utility of consumption, wages, and taxes on health, initial health, the initial employment index, and Z. If ζ_{it} follows a random walk, and innovations are uncorrelated with the initial value of health and Y_{i0}^*, our procedure should produce consistent estimates.

The expression in equation (24) can be trivially re-arranged to match our empirical specification (2), that we repeat here for reference

$$Y_{it}^* = \theta_0 + \theta_H H_{it}^* + \theta_X X_{it} + e_{it}$$

In the above equation, and as discussed in Section III, X includes all variables in Z (i.e. time dummies and an age polynomial) and the initial condition in health H_0. The only difference relative to equation (24) is that the initial index Y_{i0}^*, which is not observed, is approximated by

a function of initial employment Y_{i0} and a set of other variables that determine initial labour market attachment, including working experience accumulated so far, wealth, marital status and health in childhood.

Crucially for our purposes, the parameter of interest θ_H is the same in the two equations. Therefore, the key insight from this exercise is that, by controlling for initial health and employment while focusing on the effects of changes in health, we are able to eliminate bias in the estimation of θ_H that is induced by the potential correlation between unobserved heterogeneity in health, preferences for work and wages, (ψ_i, ξ_i, ϕ_i). As a by-product, we are also capable of revealing the response mechanisms encompassed in the parameter θ_H.

Now suppose that, instead of controlling for initial conditions, we depart from equation (21) and project the marginal value of the additional income on current health and the exogenous variables Z:

$$C_{it}^{-\gamma} W_{it}(1 - \tau_{it}) = \tilde{\delta}_0 + \tilde{\delta}_H H_{it} + \tilde{\delta}_Z Z_{it} + \widetilde{\omega}_{it}$$

Replacing in equation (21) yields

(26) $$Y_{it}^* = \tilde{\theta}_0 + \tilde{\theta}_H H_{it}^* + \tilde{\delta}_Z Z_{it} + \xi_i + \zeta_{it}$$

where $\tilde{\theta}_H = \theta_1 + \tilde{\delta}_H$. Equation (26) matches our empirical employment model without initial conditions.

The two empirical models in equations (24) and (26) differ in two substantial manners. First, the parameters identified in each case, θ_H and $\tilde{\theta}_H$, are not the same. While θ_H encompasses all the five response channels we describe in Section IV.B, $\tilde{\theta}_H$ does not account for the indirect impact on employment of the contemporaneous effects of health on productivity and benefit entitlement. And second, by not exploiting longitudinal information, model (26) does not eliminate unobserved heterogeneity correlated with current health. As a consequence, estimates of $\tilde{\theta}_H$ are likely to be biased and we would expect the direction of the bias to be positive.

50

D   Using the reduced form regressions to assess structural models of health

The findings from the reduced form model inform the structural work, just as the structural model can help us interpret the reduced form work. Our three key findings for structural modeling are as follows.

First, and most importantly, a carefully constructed single health index captures well the incentives for labor supply. We found relatively little evidence against the assumption of a single health index in our reduced form analysis, and this finding supports the use of a single index in structural models. In fact, the vast majority of life cycle models that account for health consider only a single health index (see French, 2005; French and Jones, 2011; French et al., 2018; Braun et al., 2015; De Nardi et al., 2017; Pashchenko and Porapakkarm, 2013; Aizawa and Fu, 2017, as well as the references in Footnote 24. Exceptions include Capatina et al., 2018, and Gustman and Steinmeier, 2014). However, we found that the most commonly used measure of health in structural studies, which assesses whether the respondent has a health condition that limits work, understates the impact of health on labor employment modestly in the HRS and strongly in ELSA relative to our preferred measure.

Second, dynamics are important. Accounting for initial conditions, and thus exploiting more transitory fluctuations in health, reduces the estimated impact of health by about half in England and a quarter in the US. Our model shows why this is important and reveals several channels through which changes in health affects changes in employment. We should point out that it is not obvious which of these channels are most important.

The model we described does not include all the channels by which health and employment may be related. For instance, it is conceivable that higher incomes cause better health. The Grossman (1972) model implies that those with higher income may be able to purchase better nutrition and health care, improving later health outcomes. The structural analyses of models allowing for both directions of causality is becoming increasingly common.[xxiv] Another

potential mechanism is embedded in the learning-by-doing model, whereby workers productivity on the job, and hence wages, increase with accumulated working experience.[xxv] In that case, bad health shocks that lower current employment affect future wages because of the loss in working experience. The consequent lower future wages negatively affect future employment, even if health recovers. We believe that more structural work is necessary to disentangle the various mechanisms by which employment and health are related.

Third, the US/England differences in estimates are notable. We noted an important institutional difference between the two counties in that the US disability system provides a relatively generous health benefit that is conditional on not working. Thus unhealthy Americans have a strong incentive not to work. Compared to the US, England provides more generous out of work benefits for reasons unrelated to health, such as unemployment benefits, but relatively less generous disability benefits. These institutional differences suggest that modeling the labor supply incentives of the disability insurance system is key to better understand how health affects employment decisions.

VII   Conclusions

This paper aims to provide a better understanding of the role of different measurements of health in the estimation of the impact of health on employment. We find, broadly, that estimates of the share of the decline in employment explained by declines in health are remarkably robust to the choice of health variable used; using a single subjective measure of health, multiple subjective measures, multiple objective measures, or subjective measures instrumented with objective measures makes little difference to our estimates. We conclude that this suggests measurement error and justification bias are not important sources of bias, or at least that the two sources of bias offset one another. We also find that while cognition is highly correlated with employment, including it as additional health measure does not have a dramatic impact either. These findings are consistent across the US and England.

52

We do find that our estimates are sensitive to four important modelling decisions, however. First, controlling for initial conditions such as initial health and employment considerably lowers estimates, suggesting cross sectional estimates of the relationship between health and employment are biased. Second, consistent with Bound (1991), we find that using a very small number of objective measures results much smaller estimates, suggesting these estimates suffer from omitted variable bias. Third, health is a more important driver of employment among high school dropouts, and its effects tends to drop with education. And fourth, our estimates are consistently much larger in the US than in England. This is driven predominantly by the impact of health on employment, rather than by differential declines in employment or health. It suggests that institutional setting is a key component in determining the impact health has on employment.

References

Aizawa, N. and C. Fu (2017). Heterogeneous equilibrium impacts of the affordable care act. Mimeo.

Anger, S. and G. Heineck (2010). Cognitive abilities and earnings–first evidence for Germany. *Applied Economics Letters* 17(7): 699–702.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies* 67(3): 499–527.

Banks, J., R. Crawford, and G. Tetlow (2015). Annuity choices and income drawdown: evidence from the decumulation phase of defined contribution pensions in England. *Journal of Pension Economics & Finance* 14(4): 412–438.

Banks, J., S. Keynes, and J. P. Smith (2016). Health, disability and mortaility differences at older ages between the USA and England. *Fiscal Studies*.

Banks, J., M. Marmot, Z. Oldfield, and J. P. Smith (2006). Disease and disadvantage in the United States and in England. *Journal of the American Medical Association* 295(17): 2037–2045.

Banks, J., C. O'Dea, and Z. Oldfield (2010). Cognitive function, numeracy and retirement saving trajectories. *The Economic Journal* 120: F381–F410.

Bartel, A. and P. Taubman (1979). Health and labor market success: The role of various diseases. *The Review of Economics and Statistics*, 1–8.

Benitez-Silva, H., M. Buchinsky, H. Man Chan, S. Cheidvasser, and J. Rust (2004). How large is the bias in self-reported disability? *Journal of Applied Econometrics* 19(6): 649–670.

Blau, D. M. and D. B. Gilleskie (2001). The effect of health on employment transitions of older men. In *Worker Wellbeing in a Changing Labor Market*, pp. 35–65. Emerald Group Publishing Limited.

Blau, D. M. and D. B. Gilleskie (2008). The role of retiree health insurance in the employment behavior of older men. *International Economic Review* 49(2): 475–514.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources* 26(1): 106–138.

Bound, J., M. Schoenbaum, T. R. Stinebrickner, and T. Waidmann (1999). The dynamic effects of health on the labor force transitions of older workers. *Labour Economics* 6(2): 179–202.

Bound, J., T. Stinebrickner, and T. Waidmann (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics* 156(1): 106–129.

Braun, R. A., K. A. Kopecky, and T. Koreshkova (2015). Old, sick, alone and poor: A welfare analysis of old-age social insurance programs. Mimeo, Federal Reserve Bank of Atlanta.

Brunner, E. J. (2017). Social factors and cardiovascular morbidity. *Neuroscience and Biobehavioral Reviews* 74: 260–268.

Butler, J. S., R. V. Burkhauser, J. M. Mitchell, and T. P. Pincus (1987). Measurement error in self-reported health variables. *The Review of Economics and Statistics* 69(4): 644–650.

Capatina, E. (2015, September). Life-cycle effects of health risk. *Journal of Monetary Economics* 74: 67–88.

Capatina, E., M. Keane, and S. Maruyama (2018). Health shocks and the evolution of consumption and income over the life-cycle. Mimeo.

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

Choi, S., S. Kariv, W. Mu˙ller, and D. Silverman (2014). Who is (more) rational? *American Economic Review* 104(6): 1518–50.

Cole, H. L., S. Kim, and D. Krueger (2012, November). Analyzing the effects of insuring health risks: On the trade-off between short run insurance benefits vs. long run incentive costs. Working Paper 18572, National Bureau of Economic Research.

Crossley, T. F. and S. Kennedy (2002). The reliability of self-assessed health status. *Journal of Health Economics* 21(4): 643–658.

Currie, J. and B. C. Madrian (1999). Health, health insurance and the labor market. *Handbook of Labor Economics* 3: 3309–3416.

Cutler, D. M., E. Meara, and S. Richards-Shubik (2013). Health and work capacity of older adults: Estimates and implications for Social Security policy. Working paper, Carnegie Mellon University.

Davidson, R. and J. G. MacKinnon (2003). *Econometric Theory and Methods*. New York: Oxford University Press.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature* 48(2): 424–455.

Disney, R., C. Emmerson, and M. Wakefield (2006). Ill health and retirement in britain: A panel data-based analysis. *Journal of Health Economics* 25(4): 621–649.

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2017). The lifetime costs of bad health. Mimeo. Dwyer, D. S. and O. S. Mitchell (1999). Health problems as determinants of retirement: Are self-rated measures endogenous? *Journal of Health Economics* 18(2): 173–193.

Fonseca, R., P.-C. Michaud, T. Galama, and A. Kapteyn (2009). On the rise of health spending and longevity. Rand Working Paper WR-722.

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies* 72(2): 395–427.

French, E. and J. Jones (2016). Health, health insurance, and retirement: A survey. Mimeo.

French, E. and J. B. Jones (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica* 79(3): 693–732.

French, E., H.-M. von Gaudecker, and J. B. Jones (2018). The effect of the affordable care act on the labor supply, savings, and social security of older americans.

Gilleskie, D., E. Han, and E. Norton (2017a). Disentangling the contemporaneous and dynamic effects of human and health capital on wages over the life cycle. *Review of Economic Dynamics* 25: 350–383.

Gilleskie, D. and D. Hoffman (2014). Health capital and human capital as explanations for health- related wage disparities. *Journal of Human Capital* 8(3): 235–279.

Gilleskie, D. B., E. Han, and E. C. Norton (2017b). Disentangling the contemporaneous and dynamic effects of human and health capital on wages over the life cycle. *Review of Economic Dynamics* 25: 350 – 383. Special Issue on Human Capital and Inequality.

Giustinelli, P. and M. D. Shapiro (2018). Seate: Subjective ex ante treatment effect of health on retirement. *Michigan Retirement Research Center Research Paper* (2018-382).

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy* 80(2): 223–255.

Gustman, A. and T. Steinmeier (2014). The role of health in retirement. Working Paper 19902, National Bureau of Economic Research.

Hai, R. (2015). The determinants of rising inequality in health insurance and wages.

Halliday, T. J., H. He, L. Ning, and H. Zhang (2017). Health investment over the life-cycle. *Macroeconomic Dynamics*, 1–38.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029–1054.

Heineck, G. and S. Anger (2010). The returns to cognitive abilities and personality traits in Germany. *Labour Economics* 17(3): 535–546.

Hugonnier, J., F. Pelgrin, and P. St-Amour (2012). Health and (other) asset holdings. *The Review of Economic Studies* 57(4): 779–813.

Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica* 62(2): 467–475.

Johnston, D. W., C. Propper, and M. A. Shields (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of health economics* 28(3): 540–552.

Kapteyn, A., J. P. Smith, and A. Van Soest (2007). Vignettes and self-reports of work disability in the united states and the netherlands. *American Economic Review* 97(1): 461–473.

Kim, S. (2012, October). The labor supply and welfare effects of early access to Medicare through Social Security Disability Insurance. Mimeo.

Kitao, S. (2014). A life-cycle model of unemployment and disability insurance. *Journal of Monetary Economics* 68(0): 1 – 18.

Kreider, B. and J. V. Pepper (2007). Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association* 102(478): 432–441.

Lee, L.-F. (1992). Amemiya's generalized least squares and tests of overidenfication in simultaneous equation models with qualitative or limited dependent variables. *Econometric Reviews* 11(3): 319–328.

McClellan, M. B. (1998). Health events, health insurance, and labor supply: Evidence from the health and retirement survey. In *Frontiers in the Economics of Aging*, 301–350. University of Chicago Press.

McGarry, K. (2004). Health and retirement do changes in health affect retirement expectations? *Journal of Human Resources* 39(3): 624–648.

Newey, W. (1987). Efficient estimation of limited dependent variable models with endogeneous explanatory variables. *Journal of econometrics* 36: 231–250.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics* 29(3): 229–256.

O'Donnell, O., E. van Doorslaer, and T. van Ourti (2015). *In Handbook of Income Distribution 2*, 1491–1533.

Ozkan, S. (2014). Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle. Mimeo.

Pamela, G. and M. D. Shapiro (2018). Seate: Subjective ex ante treatment effect of health on retirement. Mimeo.

Pashchenko, S. and P. Porapakkarm (2013). Work incentives of Medicaid beneficiaries and the role of asset testing. *Available at SSRN 2323775*.

Pelgrin, F. and P. St-Amour (2016, September). Life cycle responses to health insurance status. *Journal of Health Economics* 49: 76–96.

Rivers, D. and Q. H. Vuong (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics* 39(3): 347–366.

Scholz, J. K. and A. Seshadri (2016). Health and wealth in a lifecycle model. Mimeo.

Smith, J. P. (2004). Unraveling the SES: Health connection. *Population and Development Review 30*, 108–132.

Stern, S. (1989). Measuring the effect of disability on labor force participation. *The Journal of Human Resources* 24(3): 361–395.

Tables

Table 1: ELSA and HRS years and sample sizes

	ELSA		HRS	
Year	Wave	Sample Size	Wave	Sample Size
1996			3	10,215
1998			4	13,369
2000			5	11,996
2002	1	8,008	6	10,724
2004	2	6,104	7	12,126
2006	3	6,403	8	10,618
2008	4	7,426	9	9,264
2010	5	6,620	10	13,156
2012	6	6,834	11	11,805
Total		41,395	11	103.273

Notes: Sample sizes for 50–70 year olds only. Total row gives total number of observations, meaning some individuals appear multiple times.

Table 2: Objective health variables, averages by gender

Variable	ELSA		HRS	
	Men	Women	Men	Women
Cancer	0.03	0.03	0.08	0.11
Diabetes	0.09	0.06	0.19	0.17
Sight	0.02	0.02	0.04	0.05
Hearing	0.05	0.02	0.06	0.02
Blood pressure	0.30	0.26	0.50	0.50
Arthritis	0.23	0.34	0.44	0.57
Psychiatric	0.05	0.08	0.12	0.21
Lung Disease	0.04	0.04	0.08	0.10
Stroke	0.02	0.01	0.06	0.04
Heart Attack	0.03	0.01	0.02	0.01
N	18,913	22,482	44,499	58,764

Notes: Includes individuals aged 50–70. All variables are binary measures

Table 3: Subjective health variables, averages by gender

	ELSA		HRS	
Variable	Men	Women	Men	Women
Health limits activities	0.41	0.54	0.54	0.67
Self reported health	2.61	2.57	2.75	2.78
Health limits work	0.24	0.25	0.25	0.27
N	18,851	22,446	44,500	58,773

Notes: Includes individuals aged 50–70. "Health limits activities" and "Health limits work" are binary measures; "Self-reported health" is a 5-point categorical variable, where "5" is excellent

Table 4: Cognitive variables, averages by gender

Variable	ELSA		HRS	
	Men	Women	Men	Women
Immediate recall (out of 10)	5.96	6.28	5.55	6.02
Delayed recall (out of 10)	4.67	5.14	4.48	5.08
Difficulty navigating using map	0.02	0.04	0.06	0.13
Difficulty managing money	0.02	0.01	0.04	0.04
N	18,851	22,448	44,401	58,641

Notes: Includes individuals aged 50–70.

Table 5: Coefficient Estimates - Employment Regression on Cognition and Subjective Health

| | Men | | | | Women | | | |
| | ELSA | | HRS | | ELSA | | HRS | |
	No IC's [1]	IC's [2]	No IC's [3]	IC's [4]	No IC's [5]	IC's [6]	No IC's [7]	IC's [8]
Panel A: Employment on Subjective Health								
High School Dropout	0.177***	0.085***	0.194***	0.138***	0.128***	0.057***	0.161***	0.127***
	(0.004)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
High School	0.110***	0.049***	0.158***	0.106***	0.115***	0.063***	0.140***	0.109***
	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.005)	(0.002)	(0.003)
College	0.071***	0.047***	0.096***	.070***	0.068***	0.044***	0.087***	0.077***
	(0.007)	(0.007)	(0.004)	(0.004)	(0.008)	(0.008)	(0.004)	(0.005)
Panel B: Employment on Cognition								
High School Dropout	0.087***	0.013*	0.085***	0.043***	0.058***	0.013**	0.073***	0.037***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.005)	(0.005)	(.005)	(0.005)
High School	0.033***	0.011**	0.067***	0.030***	0.031***	0.007	0.061***	0.030***
	(0.006)	(0.005)	(0.003)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)
College	0.013*	0.004	0.049***	0.031***	0.019**	-0.001	0.029***	0.018***
	(0.007)	(0.008)	(0.004)	(0.005)	(0.008)	(0.008)	(0.005)	(0.005)
Panel C: Employment on Cognition and Subjective Health Cognition								
High School Dropout	0.035***	0.001	0.044***	0.029***	0.026***	0.006	0.034***	0.016***
	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
High School	0.009*	0.005	0.035***	0.017***	0.008*	0.001	0.030***	0.016***
	(0.005)	(0.005)	(0.003)	(0.003)	(0.005)	(0.005)	(0.003)	(0.003)
College	-0.001	-0.002	0.030***	0.021***	0.006	-0.006	0.014***	0.009*
	(0.007)	(0.008)	(0.004)	(.004)	(0.008)	(0.008)	(0.005)	(0.005)
Subjective Health								
High School Dropout	0.168***	0.085***	0.185***	0.135***	0.122***	0.056***	0.153***	0.125***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
High School	0.108***	0.048***	0.151***	0.104***	0.114***	0.063***	0.134***	0.106***
	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.005)	(0.002)	(0.003)
College	0.071***	0.047***	0.090***	0.067***	0.066***	0.045***	0.084***	0.076***
	(0.007)	(0.007)	(0.004)	(0.004)	(0.008)	(.008)	(0.004)	(0.005)
Sample sizes	4,692	4,692	5,777	5,777	6,957	6,957	9,199	9,199
	6,326	6,326	18,756	18,756	7,911	7,911	29,905	29,905
	3,362	3,362	9,238	9,238	2,759	2,759	9,682	9,682

Notes: All estimates include age, age squared, and wave dummies. ICs stands for initial conditions. These include the initial value of the health and cognition variables included in the regression as well as initial employment, working experience, wealth and marital status, and health in childhood. * indicates significant at 10%,** 5%, *** 1%.

Table 6: Share of Employment Decline Explained by Cognition and Subjective Health

	Men		Women	
	ELSA	HRS	ELSA	HRS
Panel A: Subjective Health				
High School Dropout	0.072***	0.123***	0.045***	0.101***
	(0.013)	(0.017)	(0.007)	(0.014)
High School	0.041***	.111***	0.047***	0.112***
	(0.008)	(0.006)	(0.009)	(0.006)
College	0.040***	0.084***	0.023***	0.077***
	(0.011)	(0.010)	(0.008)	(0.009)
Panel B: Cognition				
High School Dropout	0.009	0.037***	-0.003	0.057***
	(0.003)	(0.010)	(0.003)	(0.011)
High School	0.001	0.028***	-0.001	0.033***
	(0.002)	(0.004)	(0.001)	(0.005)
College	-0.001	0.034***	0.000	0.018***
	(0.002)	(0.007)	(0.001)	(0.006)
Panel C: Cognition and Subjective Health				
High School Dropout	0.072***	0.145***	0.043***	0.124***
	(0.013)	(0.019)	(0.008)	(0.016)
High School	0.041***	0.125***	0.047***	0.127***
	(0.008)	(0.007)	(0.009)	(0.007)
College	0.040***	0.104***	0.023***	0.085***
	(0.012)	(0.011)	(0.008)	(0.011)
Sample sizes	4,692	5,777	6,957	9,199
	6,326	18,756	7,911	29,905
	3,362	9,238	2,759	9,682

*Notes: All estimates include age, age squared, wave dummies and the full set of initial conditions. Standard errors are bootstrapped with 500 repetitions. * indicates significant at 10%, ** 5%, *** 1%.*

Table 7: Percent Differences in the Explained Share of Employment Decline and *p*-values for Testing Null of No Differences – Explanatory Value of Adding Cognition

	Percent differences				p-values			
	Men		Women		Men		Women	
	ELSA	HRS	ELSA	HRS	ELSA	HRS	ELSA	HRS
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
				Panels C versus A of Table 6				
High School Dropout	0.1	18.1	-4.3	22.1	0.499	0.189	0.439	0.146
High School	-0.3	12.7	-0.4	13.4	0.496	0.078	0.495	0.055
College	1.6	22.6	3.2	10.2	0.484	0.111	0.473	0.298

Notes: Estimates of relative differences in Columns 1–4 compare figures in Panels A and C of Table 6, with Panel A as the baseline. p−values in Columns 5–8 for testing the equality of the same δ estimates.

Table 8: Share of Employment Decline Explained by Subjective Health - Various Specifications

	Men		Women	
	ELSA	HRS	ELSA	HRS
Panel A: First principal component				
High School Dropout	0.072***	0.123***	0.045***	0.101***
	(0.013)	(0.017)	(0.007)	(0.014)
High School	0.041***	0.111***	0.047***	0.112***
	(0.008)	(0.006)	(0.009)	(0.006)
College	0.040***	0.084***	0.023***	0.077***
	(0.011)	(0.010)	(0.008)	(0.009)
Panel B: Three subjective measures separately				
High School Dropout	0.060***	0.106***	0.033***	0.107***
	(0.015)	(0.017)	(0.009)	(0.017)
High School	0.032***	0.110***	0.039***	0.123***
	(0.011)	(0.007)	(0.008)	(0.007)
College	0.011	0.080***	0.028***	0.078***
	(0.013)	(0.012)	(0.011)	(0.010)
Panel C: Health limits work				
High School Dropout	0.029***	0.084***	0.014*	0.104***
	(0.011)	(0.016)	(0.007)	(0.016)
High School	0.020***	0.101***	0.020***	0.119***
	(0.006)	(0.007)	(0.006)	(0.006)
College	-0.003	0.068***	0.003	0.071***
	(0.007)	(0.009)	(0.006)	(0.009)
Sample sizes	4,692	5,777	6,957	9,199
	6,326	18,756	7,911	29,905
	3,362	9,238	2,759	9,682

Notes: All estimates include age, age squared, wave dummies and the full set of initial conditions. Standard errors are bootstrapped with 500 repetitions. * indicates significant at 10%,** 5%, *** 1%.

Table 9: Percent Differences in the Explained Share of Employment Decline and *p*-values for Testing Null of No Differences – Explanatory Value of Added Subjective Health Information

	Percent differences				p-values			
	Men		Women		Men		Women	
	ELSA	HRS	ELSA	HRS	ELSA	HRS	ELSA	HRS
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	Panels B vs A (three separate subjective measures)							
High School Dropout	-16.0	-13.7	-27.5	5.4	0.076	0.039	0.075	0.274
High School	-21.1	-1.0	-16.9	9.8	0.148	0.404	0.171	0.001
College	-72.8	-5.4	26.4	1.5	0.001	0.183	0.231	0.411
	Panels C vs A (health limits work)							
High School Dropout	-58.9	-32.1	-68.6	3.0	0.000	0.002	0.000	0.388
High School	-51.6	-8.9	-58.4	6.1	0.000	0.048	0.001	0.084
College	-106.3	-19.6	-84.6	-8.1	0.000	0.012	0.005	0.193

Notes: Estimates of relative differences in Columns 1–4 compare figures in Panels A to C of Table 8, with Panel A as the baseline. p−values in Columns 5–8 for testing the equality of the same δ estimates.

Table 10: Share of Employment Decline Explained by Subjective Health and Cognition – Subjective Health Instrumented using Objective Health

| | Men | | Women | |
	ELSA	HRS	ELSA	HRS
Panel A: Subjective health				
High School Dropout	0.086***	0.142***	0.055***	0.136***
	(0.022)	(0.024)	(0.015)	(0.023)
High School	0.053***	0.112***	0.058***	0.134***
	(0.015)	(0.011)	(0.016)	(0.011)
College	0.052***	0.132***	0.028**	0.100***
	(0.019)	(0.021)	(0.013)	(0.018)
Panel B: Subjective health and cognition				
High School Dropout	0.085***	0.158***	0.054***	0.147***
	(0.022)	(0.024)	(0.016)	(0.022)
High School	0.053***	0.122***	0.058***	0.144***
	(0.014)	(0.011)	(0.016)	(0.011)
College	0.054***	0.142***	0.029**	0.103***
	(0.020)	(0.020)	(0.014)	(0.018)
Sample sizes	4,692	5,777	6,957	9,199
	6,326	18,756	7,911	29,905
	3,362	9,238	2,759	9,682

Notes: All estimates include age, age squared, wave dummies and the full set of initial conditions. Standard errors are bootstrapped with 500 repetitions. * indicates significant at 10%,** 5%, *** 1%.

Table 11: Percent Differences in the Explained Share of Employment Decline and p-values for Testing Null of No Differences – Comparing OLS and IV estimates

	Percent differences				p-values			
	Men		Women		Men		Women	
	ELSA	HRS	ELSA	HRS	ELSA	HRS	ELSA	HRS
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panels B vs A (three separate subjective measures)								
High School Dropout	19.6	15.4	22.5	34.5	0.294	0.272	0.280	0.100
High School	29.8	1.1	22.5	20.1	0.247	0.462	0.284	0.048
College	31.2	55.7	23.0	29.4	0.295	0.021	0.377	0.136
Panels C vs A (health limits work)								
High School Dropout	18.8	9.0	27.0	18.8	0.300	0.342	0.263	0.198
High School	29.6	-2.2	23.2	13.1	0.247	0.415	0.275	0.117
College	32.9	37.3	22.7	21.1	0.290	0.049	0.378	0.201

Notes: Estimates of relative differences in Columns 1–4 compare figures in Panels A and C of Table 6 with those in Panels A and B of Table 10, with Table 6 as the baseline. p−values in Columns 5–8 for testing the equality of the same δ estimates.

Table 12: Overidentification Test

	Men		Women	
	ELSA [1]	HRS [2]	ELSA [3]	HRS [4]
Panel A: Subjective Health				
High School Dropout	0.221	0.217	0.134	0.001
High School	0.106	0.000	0.284	0.000
College	0.280	0.000	0.093	0.000
Panel B: Subjective Health, with Cognition				
High School Dropout	0.203	0.238	0.136	0.001
High School	0.110	0.000	0.290	0.000
College	0.283	0.000	0.079	0.000

Notes: Table compares F-Statistic to χ^2 Critical Values, giving $p-$values for the null of no statistical relationship between our objective measures and the IV residuals.

Table 13: Share of Employment Decline Explained by Objective Health

	Men		Women	
	ELSA	HRS	ELSA	HRS
Panel A: Blood pressure only				
High School Dropout	0.030**	0.050***	0.013	0.053***
	(0.013)	(0.013)	(0.009)	(0.012)
High School	0.008	0.023***	0.004	0.024***
	(0.009)	(0.006)	(0.007)	(0.006)
College	0.007	0.034***	0.005	0.035***
	(0.009)	(0.010)	(0.010)	(0.009)
Panel B: Add Arthritis, Psychiatric, Lung				
High School Dropout	0.061***	0.102***	0.045***	0.126***
	(0.017)	(0.015)	(0.015)	(0.017)
High School	0.022*	0.061***	0.039***	0.075***
	(0.012)	(0.009)	(0.013)	(0.009)
College	0.024*	0.062***	0.004	0.056***
	(0.013)	(0.014)	(0.016)	(0.015)
Panel C: Add Cancer, Diabetes, Stroke, Heart Attack				
High School Dropout	0.080***	0.156***	0.068***	0.181***
	(0.018)	(0.019)	(0.017)	(0.022)
High School	0.033**	0.081***	0.039***	0.098***
	(0.013)	(0.011)	(0.013)	(0.010)
College	0.038**	0.096***	0.012	0.075***
	(0.015)	(0.017)	(0.016)	(0.016)
Panel D: Add Sight, Hearing - full specification				
High School Dropout	0.081***	0.152***	0.067***	0.197***
	(0.019)	(0.019)	(0.018)	(0.023)
High School	0.034**	0.084***	0.038***	0.101***
	(0.013)	(0.011)	(0.013)	(0.010)
College	0.038**	0.100***	0.017	0.073***
	(0.015)	(0.017)	(0.016)	(0.017)
Sample sizes	4,692	5,777	6,957	9,199
	6,326	18,756	7,911	29,905
	3,362	9,238	2,759	9,682

*Notes: All estimates include age, age squared, wave dummies and the full set of initial conditions. Standard errors are bootstrapped with 500 repetitions. * indicates significant at 10%,** 5%, *** 1%.*

Table 14: Percent Differences in the Explained Share of Employment Decline and *p*-values for Testing Null of No Differences – Comparing Subjective and Objective Health Measures

| | Percent differences | | | | p-values | | | |
| | Men | | Women | | Men | | Women | |
	ELSA [1]	HRS [2]	ELSA [3]	HRS [4]	ELSA [5]	HRS [6]	ELSA [7]	HRS [8]
	Panel D of Table 13 vs Panel A of Table 10							
High School Dropout	-5.7	6.8	21.6	44.6	0.405	0.331	0.163	0.006
High School	-36.6	-24.9	-33.9	-25.0	0.022	0.001	0.066	0.000
College	-27.4	-23.9	-38.8	-26.4	0.159	0.028	0.159	0.004

Notes: Estimates of relative differences in Columns 1–4 compare figures in Panel D of Table 13 and those in Panel A of Table 10, with the latter as baseline. p−values in Columns 5–8 for testing the equality of the same δ estimates.

Table 15: Oaxaca Decomposition of US-English differences

	Men			Women		
	θ	ΔH	ΔY	θ	ΔH	ΔY
Subjective health						
High School Dropout	0.97	-0.04	0.07	0.85	-0.24	0.38
High School	0.97	-0.08	0.11	0.69	0.09	0.22
College	0.57	0.11	0.32	0.76	0.03	0.20
Subjective health and cognition						
High School Dropout	1.03	-0.08	0.05	0.85	-0.19	0.33
High School	0.93	-0.02	0.10	0.74	0.06	0.20
College	0.70	0.01	0.30	0.86	-0.06	0.20

Notes: Decomposition of the US-English differences in the estimates of δ by its different components. Estimates are blanked out where they are uninformative. The columns labelled 'θ', 'ΔH' and 'ΔY' show the shares explained by differences in the estimated coefficients, health declines and employment declines, respectively.

Figures
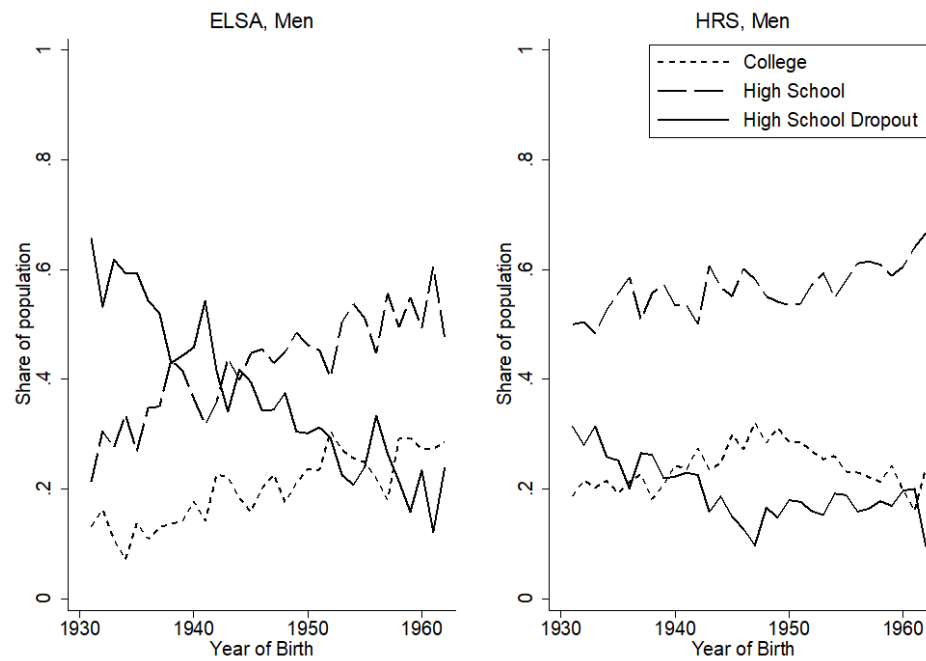
Figure 1: ELSA and HRS Education groups on D.O.B. year for men
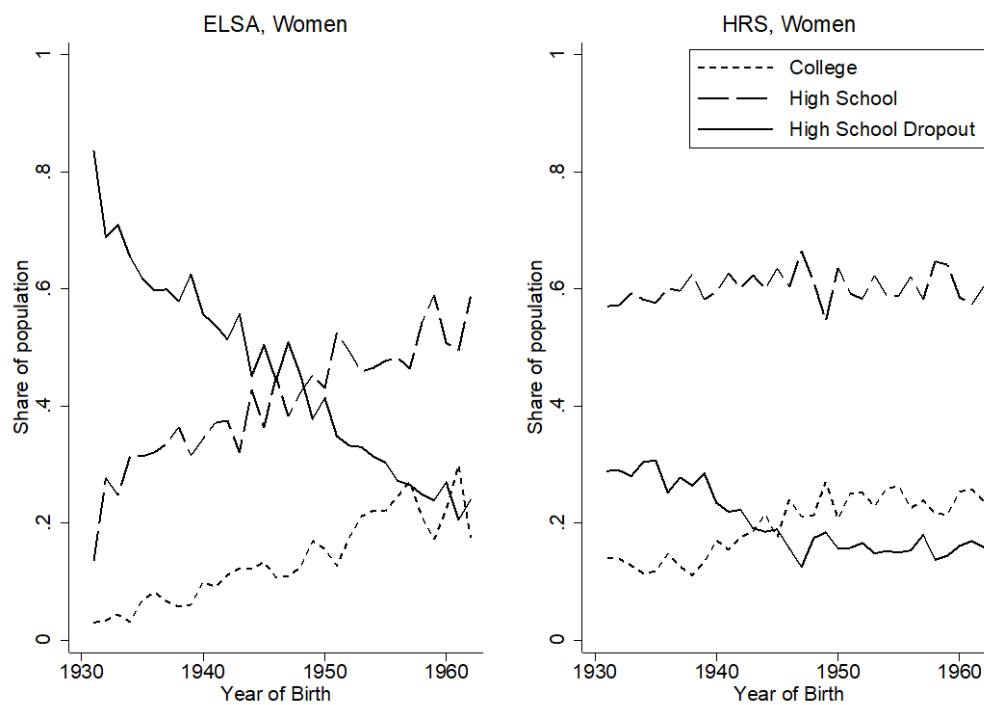
Figure 2: ELSA and HRS Education groups on D.O.B. year for women
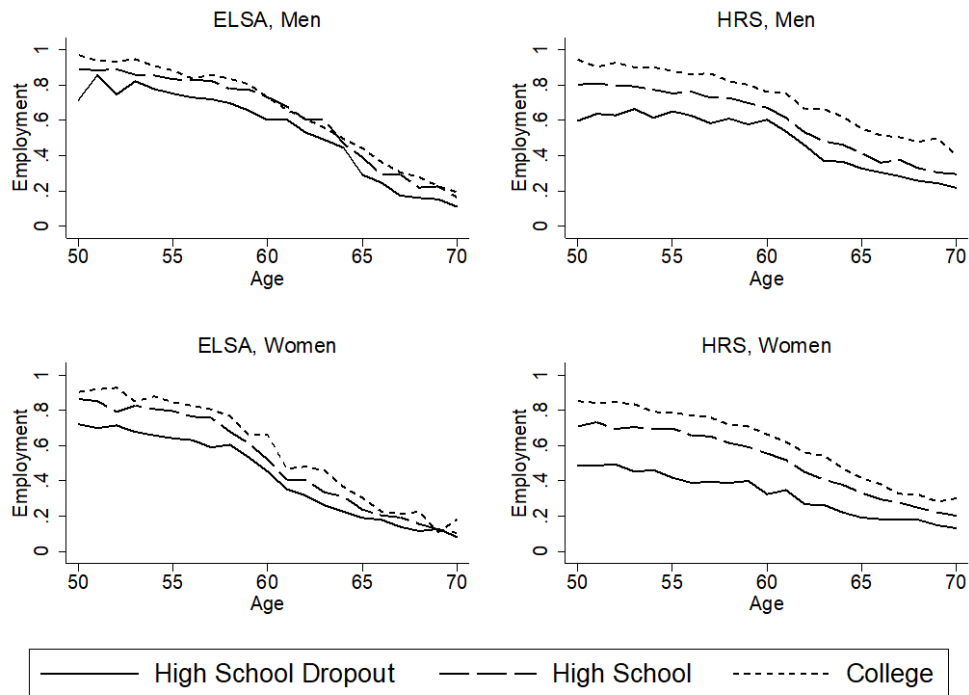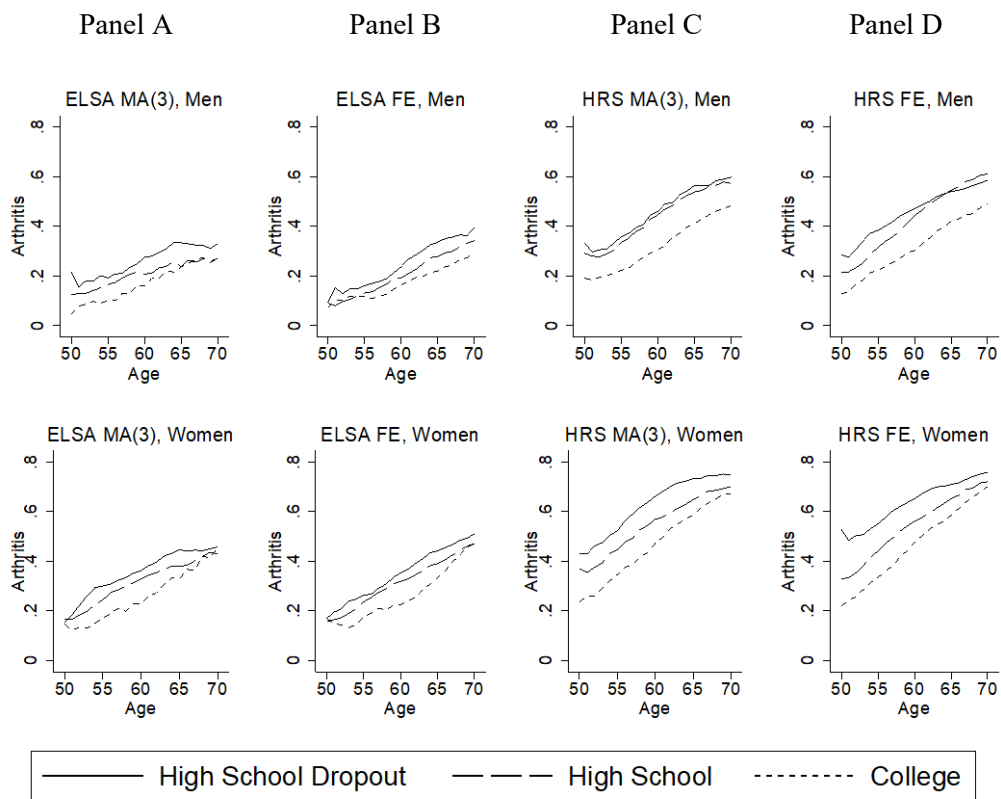
Figure 3: ELSA Employment on age, by gender and education

Figure 4: Prevalence of arthritis by age, gender and education

Notes: MA(3) indicates a 3-year moving average and FE indicates fixed effects estimates.
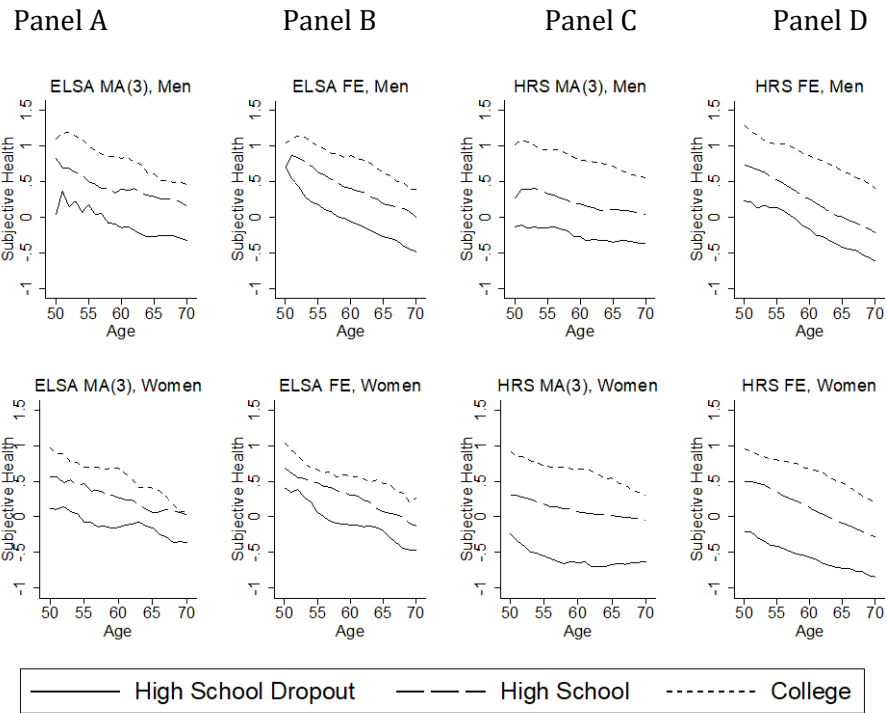
Figure 5: Single subjective health index by age, gender and education

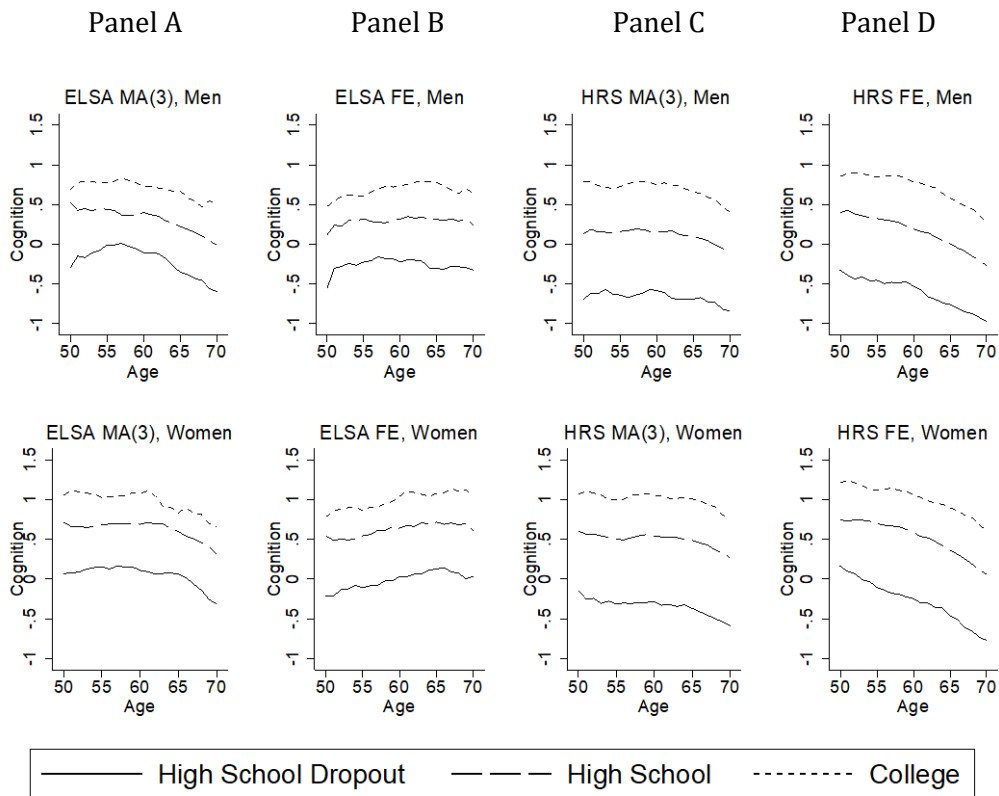*Notes: MA(3) indicates a 3-year moving average and FE indicates fixed effects estimates.*

Figure 6: Cognition index by age, gender and education

Notes: MA(3) indicates a 3-year moving average and FE indicates fixed effects estimates.

[i] Currie and Madrian (1999) state that 'although the question of how health affects participation has been intensively studied little consensus on the magnitude of the effects has been reached.' They argue that one key reason for this is the range of different approaches for measuring health. Table 4 of their paper highlights the range of estimates. Tables 18.3 and 18.4 of O'Donnell et al. (2015) highlight the same qualitative findings hold from the more recent literature addressing this question. For example, they show that French (2005) estimates that a work limiting physical impairment or nervous condition results in a 45ppt reduction in the probability of employment at age 62, while Smith (2004) estimates that a new major diagnosis is associated with a 15ppt reduction for 50-62 year olds. Smith (2004) also estimates much small effects for minor diagnoses, aligning with McClellan (1998).

[ii] This is an innocuous standardization to ensure that all health variables are measured on a similar scale, that of H.

[iii] See Online Appendix Section 4.1. There is some ambiguity as to whether it is appropriate to include these ADL measures as objective health measures, but we decided to follow the common practice and exclude them.

[iv] We also used Factor Analysis, and obtained results that were very similar to those we report here. The measures of subjective health and, more broadly, the datasets we use in the empirical exercise are described in Section IV below.

[v] While it would also be possible to construct an index of health based on the objective variables, it would not be as compelling to do so as objective measures reflect different aspects of health, rather than the same latent index.

[vi] Not excluding the Second and Third Principal Components means rejecting the joint hypotheses of a single index, model specification (such as linearity, homogeneity, etc.) and no measurement error. However, not rejecting the joint hypotheses shows that the single index assumption is difficult to reject.

[vii] Although failure to reject the null supports the single index assumption, the results from this test should be considered cautiously. As noticed by Deaton (2010) the exclusion restrictions are an IV identification assumption that cannot be tested, even in the presence of multiple instruments. In our case, the residuals $\hat{e}^{IV}$ can be orthogonal to the instruments even if the single index assumption does not hold, because in such case orthogonality is being tested at a biased estimate of θ_H (Newey, 1985). In turn, in cases where the single index assumption is valid but the impact of health is heterogeneous, each instrument may be valid in isolation (identifying effects at different margins, for different sub-populations). But by taking all instruments together it may be impossible to find a value of $\hat{\theta}_1^{IV}$ for which the orthogonality conditions are satisfied (Imbens and Angrist, 1994; Angrist et al., 2000).

[viii] As for health, we investigate the use of Factor Analysis as an alternative but find almost no difference in the results.

[ix] As noted before and in Bound (1991), estimates of the coefficients associated with other drivers of employment X may be biased even when $\hat{\theta}_M$ is not. Note that the only parameter we use to calculate δ_M in the linear framework (equation 1) is $\hat{\theta}_M$, so predictions from a linear model will not be affected by bias in other coefficients. However, the marginal effects in the non-linear model depend on all parameters and, hence, may be affected.

[x] These groupings closely resemble those used in Banks et al. (2015).

[xi] Both datasets also provide information on working hours and hourly wages. Considering working hours instead of the dichotomous employment outcome does not change our findings, so we omit it here. Results for hourly wage rates, however, were much nosier than those for employment. This was not unexpected as selection into work is likely to play a key role in determining estimates of the impact of health on hourly wages if those who remain in work are healthier than those who drop out (and increasingly so with age). The age profiles of hourly wages and working hours can be found in the Online Appendix, but we do not further investigate these impacts here.

[xii] Plots for the each of the subjective measures can be found in Section 2.3 of the Online Appendix, while more detail on the distribution of the measures and the weights assigned to each variable and the estimates from the first stage IV regression can all be found in Section 3.1 of the Online Appendix.

[xiii] See Banks et. al (2010) for a good description of the cognitive function measures in ELSA and Choi et. al (2014) for more on measures of cognition and how they vary with age, gender and education.

[xiv] ELSA does include a numeracy test in some waves (specifically, waves 1, 4 and 6), which might be considered a crystalized measure (and is used in Banks et. al, 2010).

[xv] Plots for each of the component variables are given in Section 2.6 of the Online Appendix, while the weights assigned to each variable can be found in Section 3.1 of the Online Appendix.

[xvi] We found little evidence that these results are being driven by learning of the tests, which we investigated by removing the first wave individuals were surveyed, with the idea that the majority of learning should occur between the first and second waves individuals are observed. These figures are available from the authors on request.

[xvii] An intermediate specification including the two first principal components was also tried. It showed very similar results to those in Panel B. These are available from the authors upon request.

[xviii] Attenuation bias from measurement error is a more serious problem when using the subjective measures separately (as in Panel B of Table 8) than for estimates based on the single composite subjective health index (as in Panel A). This is because measurement error that is *not* common across the underlying subjective health measures is cleared from the index but will contaminate estimates based directly on the observed subjective variables. This can help explaining why some of the estimates in Panel B are *lower* than their counterparts in Panel A of the Table.

[xix] In practice we do the non-linear version of this test.

[xx] We also estimated the effects of health on employment using each of the objective measures on their own, and then every pair combination. Consistently with results in Panel A and those of Bound (1991) before us, the estimates obtained in this way are always very small and mostly statistically insignificant at conventional levels. Moreover, we reject the hypothesis of equality between these effects and those obtained using the full specification in Panel D in the vast majority of cases. The equality between the effects based on two objective conditions and those obtained using the full specification is not rejected only in 7 out of 435 cases. Using a single condition, we reject the null of

equality in all cases (out of 110). This clearly shows that very parsimonious models lead to systematic downward bias in measuring the impact of health on employment.

[xxi] A description of the decomposition procedure can be found in Section 5 of the Online Appendix.

[xxii] These are only some of the mechanisms driving employment changes among older workers. They may also face increasingly unfavorable incentives to work created by the tax, benefit and pension systems. Changes in health interact with these other mechanisms by altering the value of wealth holdings and the entitlement to pensions and benefits.

[xxiii] One mechanism that we do not consider explicitly is medical expenses. This is important in the US (see Pashchenko and Porapakkarm, 2013; Kitao, 2014; Kim 2012) but less so in the UK or in most other European countries, where full coverage of medical expenditures is independent of income and employment.

[xxiv] See Ozkan (2014), Fonseca et al. (2009), Blau and Gilleskie (2008), Pelgrin and St-Amour (2016), Cole et al. (2012), Hai (2015), Halliday et al. (2017), Hugonnier et al. (2012), and Scholz and Seshadri (2016). Outside the economics field, the predominant view is indeed that income causes health rather than vice-versa (see Brunner, 2017, for a recent review).

[xxv] 25Examples of papers that account for this mechanism include Capatina et al. (2018) and Gilleskie et al. (2017b).