

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**

August 2017

Abstract

Estimates of effect of health on employment differ from study to study due to differences in methods, data, institutional background and health measure. We assess the importance of these differences, using a unified framework to interpret and contrast estimate for the US and England. We find that subjective and objective health measures, and subjective measures instrumented by objective measures produce similar estimates but only if a sufficiently large number of objective measures is used. Otherwise, objective measures produce downward biased estimates. Failure to account for initial conditions produces upward biased estimates. We find that a single subjective health index yields similar estimates to multiple measures. Overall, declines in health explain up to 15% of the decline in employment between ages 50 and 70. The effects drop with education and are larger in the US than in England. Cognition has little added explanatory power once we control for health. *I10, J24, J26, E24*

Keywords: Health, Cognition, Labor Supply, Retirement

*Corresponding author: Monica Costa Dias; contact: monica_d@ifs.org.uk, 7 Ridgmount Street, London WC1E 7AE, UK. We thank Robert Willis, Naoki Aizawa, and seminar participants at the Conference on Working Longer at IFS and the ELSA Wave 7 Launch Conference for comments. Financial support from the Michigan Retirement Research Center (MRRC), Grant G-2014-13567 from the Alfred P. Sloan Foundation, and the ESRC Centre for the Microeconomic Analysis of Public Policy at the IFS is gratefully acknowledged. Jack Britton also thanks the British Academy for financial support. The views expressed in this paper are those of the authors and not necessarily those of the Social Security Administration, the MRRC or the British Academy.

1 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.

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 index is not readily observable. This lead 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 favor ‘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. Even within the objective and subjective categories, there is no agreement about which specific variables should be used. Moreover, various modeling strategies have also been adopted, often resulting in different health effect parameters. 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? How should these health measures 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 account for 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 an instrument for the subjective measures. We use principal components and factor analysis to combine multiple health measures into a parsimonious single health index. 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 health impairs work.

Our empirical analysis is based on two large 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 up to about half. 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 2 to 3. 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 2 provides an overview of the literature investigating the impact of health on labor supply. Section 3 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 4 describes the ELSA and HRS datasets and our constructed measures of health and cognition. Section 5 presents our main estimates and examines the sources of differences between the US and England. Section 6 concludes.

2 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. 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 (see also French (2005)).

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) and McGarry (2004) circumvent concerns of justification bias by examining the relationship between health and *expected* retirement. 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. The structural analyses of models allowing for both is becoming increasingly common.¹ Outside the economics field, the predominant view is indeed that income causes health rather than vice-versa (see Brunner (2017) for a recent review). On the other hand, the simultaneous determination of health and employment could result from common (unobserved) drivers of both

¹ 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).

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) 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. 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. (2016), Braun et al. (2015), De Nardi et al. (2017), Pashchenko and Porapakarm (2013), Aizawa and Fu (2017), as well as the references in Footnote 1. An exception is Gustman and Steinmeier (2014)). But whether the single index is a sufficiently detailed representation of health remains an open question. We show that a single health index captures the variation in health well. To the best of our knowledge, we are the first to test the single index assumption.

3 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 the 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

(H), but this is not readily observed in the data. Current data sets do not include all the health variables that affect work capacity, and those that are included may suffer from measurement error and justification bias; alternative estimation approaches deal differently with these problems. Here we show how 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? Here we show how to use multiple objective and subjective measures to answer these questions.

In what follows, we discuss the estimation of the following simple model of employment. For individual i at time t :

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

where Y is employment, H is health status and X are other drivers of employment, which include a second order age polynomial, marital status, and time dummies. In X we also include initial conditions in health and employment, measured when each respondent is first observed in the sample, and accumulated years of work. This is critical to deal with potential bias from common unobserved factors driving both employment and health. Conditionally on X , we therefore assume that the health status H is independent of the unexplained component of employment, e . Note that this specification implicitly assumes homogeneous effects of health on employment; in particular, it implies that the impact of health is linear, so that the impact of a small change in health is independent of the existing level of health.² We will relax this assumption by considering a non-linear model of employment and show that our empirical results remain unaltered.

²In practice we estimate all parameters separately by gender and education, so homogeneity is assumed within-group.

3.1 Measuring health using objective measures

The health stock can be formalised by a combination of all health conditions (and combinations of conditions) that limit work, h_k^o for $k = 1, \dots, K$. These are typically labelled ‘objective’ health measures because they represent medical health conditions that can be unambiguously named; 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

$$H_{it} = \sum_{k=1}^K \alpha_k h_{kit}^o \quad (2)$$

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

$$Y_{it} = \theta_0 + \sum_{k=1}^K \tilde{\theta}_{Hk} h_{kit}^o + \theta_X X_{it} + e_{it} \quad (3)$$

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

In practice, the simple specification in equation (3) 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. Consequently, the effect of health can only partly be determined. Second, not all health conditions are equally important for employment, a fact that is expressed by the multiple parameters $\tilde{\theta}_{Hk}$. While some conditions may be so debilitating as to impair work at least temporarily (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, in most cases (and certainly when dealing with survey data) health information only describes whether respondents suffer from certain conditions, not how serious or limiting such condition may be. This is a key source of measurement error that is expected to bias the estimated effects towards zero.

To put it more formally, suppose that the true health stock H is a combination of two conditions, (h_1^o, h_2^o) , and for simplicity to highlight ideas we will ignore the correlation between health and the X variables. Assume we normalize the variance of the objective measures to equal that of H and ensure that all variables are ordered in the same direction (say, higher values for better health) so that $(\alpha_1, \alpha_2) \in [0, 1]^2$. Suppose that in a specific study only h_1^o is observed and that it is measured without error. In such case, the OLS estimator of θ_H yields

$$\begin{aligned} \text{plim } \hat{\theta}_H^o &= \frac{\text{Cov}(Y, h_1^o)}{\text{Var}(h_1^o)} \\ &= \frac{\text{Cov}(\theta_0 + \tilde{\theta}_{H1}h_1^o + \tilde{\theta}_{H2}h_2^o + \theta_X X + e, h_1^o)}{\text{Var}(h_1^o)} \\ &= \theta_H \alpha_1 + \theta_H \alpha_2 \frac{\text{Cov}(h_1^o, h_2^o)}{\text{Var}(H)}. \end{aligned}$$

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 condition 1, 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 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 larger than under the case where they are uncorrelated, lessening the impact of the bias. A prediction based on model estimates of how much changes in health status drives employment (as described below in Section 3.6) 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 to 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 10 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.³

3.2 Measuring health using subjective measures

Although we cannot observe H directly, we do observe the ‘subjective’ measures h_k^s . These are self-reported health measures that describe overall health status and provide an alternative to using objective measures to describe health. 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 (2)), the subjective measures are overall (noisy) measures of the single latent health stock. This idea can be formalised by a set of relations

$$h_{kit}^s = \beta_k H_{it} + u_{kit} \text{ for } k = 1, \dots, K^s \quad (4)$$

where the unobserved health stock H is the common latent factor driving all K^s subjective measures of health and 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 one health index either by Principal Component Analysis or by Factor Analysis.⁴

³Results available from the authors. There is some ambiguity as to whether it is appropriate to include these ADL measures as objective health measures. We decided to follow the common practice and exclude them.

⁴The measures of subjective health and, more broadly, the datasets we use in the empirical exercise are described

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.

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

$$H_{it}^s = H_{it} + v_{it}.$$

If e from equation (1) and the measurement error v are uncorrelated, estimates of the health effect θ_1 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 using H^s to proxy H has asymptotic limit:

$$\text{plim } \hat{\theta}_H^s = \frac{\theta_H \text{Var}(H) + \text{Cov}(e, v)}{\text{Var}(H) + \text{Var}(v)} \quad (5)$$

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.

3.3 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, while approaches using only subjective measures suffer from measurement error and

in Section 4 below.

justification bias. 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 coefficient (call it η) converges in probability to

$$\begin{aligned} \text{plim } \hat{\eta} &= \frac{\text{Cov}(H^s, h_k^o)}{\text{Var}(h_k^o)} \\ &= \frac{\text{Cov}(H, h_k^o)}{\text{Var}(H)} \\ &= \frac{\alpha_k \text{Var}(H) + \sum_{l \neq k} \alpha_l \text{Cov}(h_l^o, h_k^o)}{\text{Var}(H)} \end{aligned}$$

Recall that H is a combination of all objective health conditions (as described in equation (2)), which have 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 is

$$\begin{aligned} \text{plim } \hat{\theta}_H^{IV} &= \frac{\text{Cov}(Y, \eta h_k^o)}{\eta^2 \text{Var}(h_k^o)} \\ &= \theta_H \frac{\text{Cov}(H, h_k^o)}{\eta \text{Var}(H)} = \theta_H. \end{aligned}$$

Under the IV exclusion restrictions, we can assess the importance of biases confounding estimates of θ in model (1) when estimates are based on objective measures (due to omitted variables) and 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.

3.4 Tests of the single index assumption

We now turn to discuss the plausibility of the single index assumption. Explicitly, we state the ‘single index assumption’ as follows: the subjective health index H^s , constructed as a composite

measure of the variation in the subjective health variables, contains all relevant health information for employment. This is a stronger assumption than that implicit in model (1), which implies that a single dimension of health (H) captures all the variation in health-related work capacity. In model (1), H can be a function of multiple health conditions, with varying implications for work capacity as described in equation (2). In contrast, our single index assumption requires that a summary of multiple measures of self-reported health status, which are not themselves necessarily related to work capacity, captures all health-related variation in work capacity. Notice, however, that measurement error and justification bias are not ruled out by the single index assumption. Indeed, we do allow for both sources of noise in H^s , as described before.

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.⁵

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

⁵Not 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.

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

A slightly more subtle point is that the IV approach with multiple instruments provides the means to test the validity of the single index assumption using a Sargan over-identification 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 10 objective measures, we have 9 over-identification conditions.

In practice, we implement the test following the suggestion in Davidson and MacKinnon (2003). We construct the IV residuals:

$$\hat{e}_{it}^{IV} = Y_{it} - \hat{\theta}_0^{IV} - \hat{\theta}_H^{IV} H_{it}^s - \hat{\theta}_X^{IV} X_{it}. \quad (6)$$

Under the single index assumption, we know that:

$$E[\hat{e}_{it}^{IV} h_{kit}^o | X_{it}] = 0 \text{ for } k = 1, \dots, K^o. \quad (7)$$

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.⁶

3.5 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

⁶Although 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)).

Section 4.4 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. The extended model is

$$Y_{it} = \theta_0 + \theta_H H_{it} + \theta_C C_{it} + \theta_X X_{it} + e_{it}. \quad (8)$$

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 (again, we investigate the use of Factor Analysis as an alternative but find almost no difference in the results).

3.6 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 years period that span from 50 to 70 years of age. The parameter we calculate is

$$\hat{\delta}_Z = \frac{\hat{\theta}_Z (\bar{Z}_{70} - \bar{Z}_{50})}{\bar{Y}_{70} - \bar{Y}_{50}} \text{ for } Z = H, C$$

where the upper bar represents average predictions from a simple fixed effects regressions of health, cognition and employment on age. 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 calculate the single impact parameter

$$\hat{\delta} = \sum_j \frac{\hat{\theta}_j (\bar{Z}_{j,70} - \bar{Z}_{j,50})}{\bar{Y}_{70} - \bar{Y}_{50}} \quad (9)$$

where j indexes the various health and cognition measures included in the employment regression model. A similar metric has been used by French (2005). Cutler et al. (2013) calculate the decline in employment not explained by declining health.

4 Data and descriptive statistics

This paper uses waves 1 to 6 of the English Longitudinal Study of Ageing (ELSA), covering years 2002-2012, 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 2002/03. Later interviews were conducted in 2004/05, 2006/07, 2008/09, 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 \times time observations.

Table 1: ELSA and HRS years and sample sizes

Year	ELSA		HRS	
	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		103,273

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

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).⁷ 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. Indeed, 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 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, as initial

⁷These groupings closely resemble those used in Banks et al. (2015).

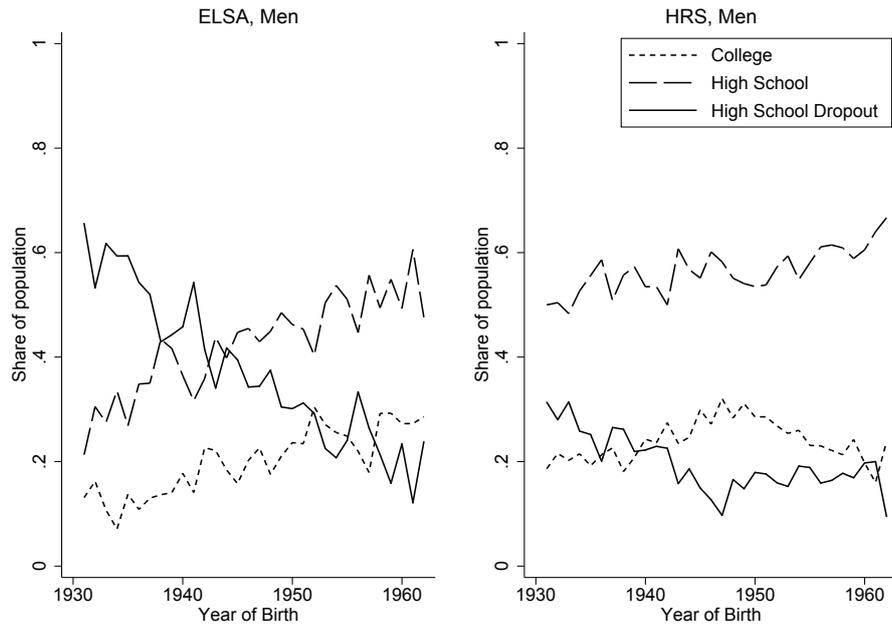


Figure 1: ELSA and HRS Education groups on D.O.B. year for men

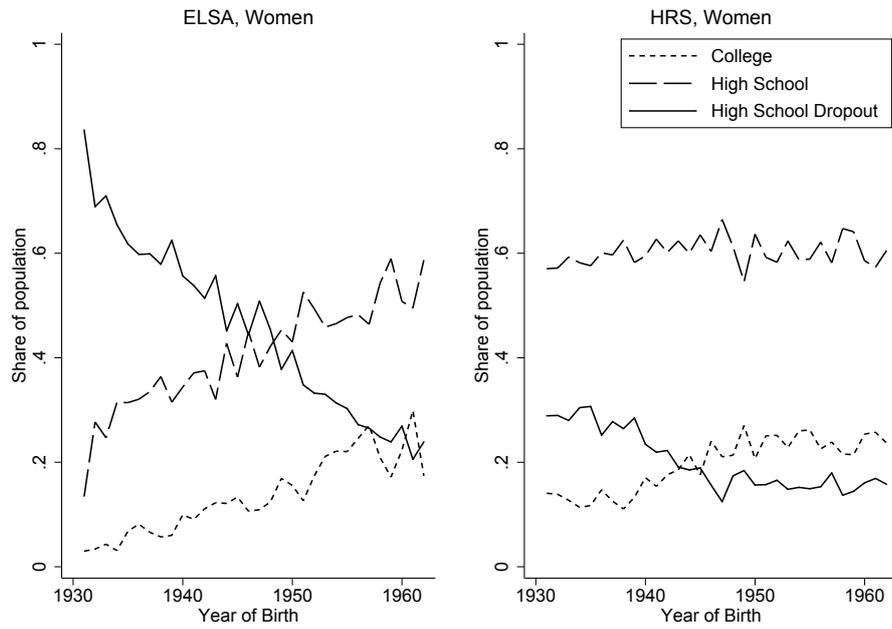


Figure 2: ELSA and HRS Education groups on D.O.B. year for women.

conditions in our regressions we use historical data on health during childhood and years of working experience to capture long-term health status and labor market attachment.

4.1 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.⁸

4.2 Objective measures of health

As described in the methods Section 3, we consider health variables in two broad categories, objective and subjective. Here we focus on the former. Table 2 summarises 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 of the 10 conditions for which the respondent is treated (top 10 rows in the Table)), and is often twice or even three times larger in magnitude. For example, cancer prevalence is 3% in

⁸Both 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 noisier 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 Section 1.1 in the Online Appendix, but we do not further investigate these impacts here.

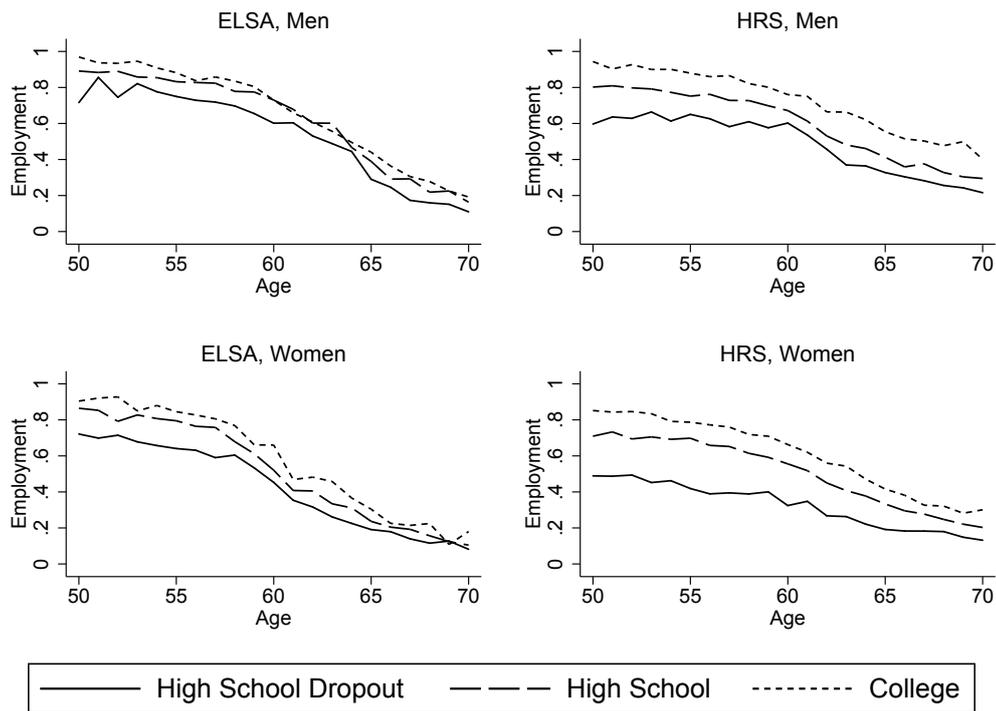


Figure 3: ELSA Employment on age, by gender and education

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

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

Includes individuals aged 50-70. All variables are binary measures.

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 fixed effects (normalised to have mean zero in the population), and β are a full set of age dummy variables that capture health-age profiles net of fixed effects. 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

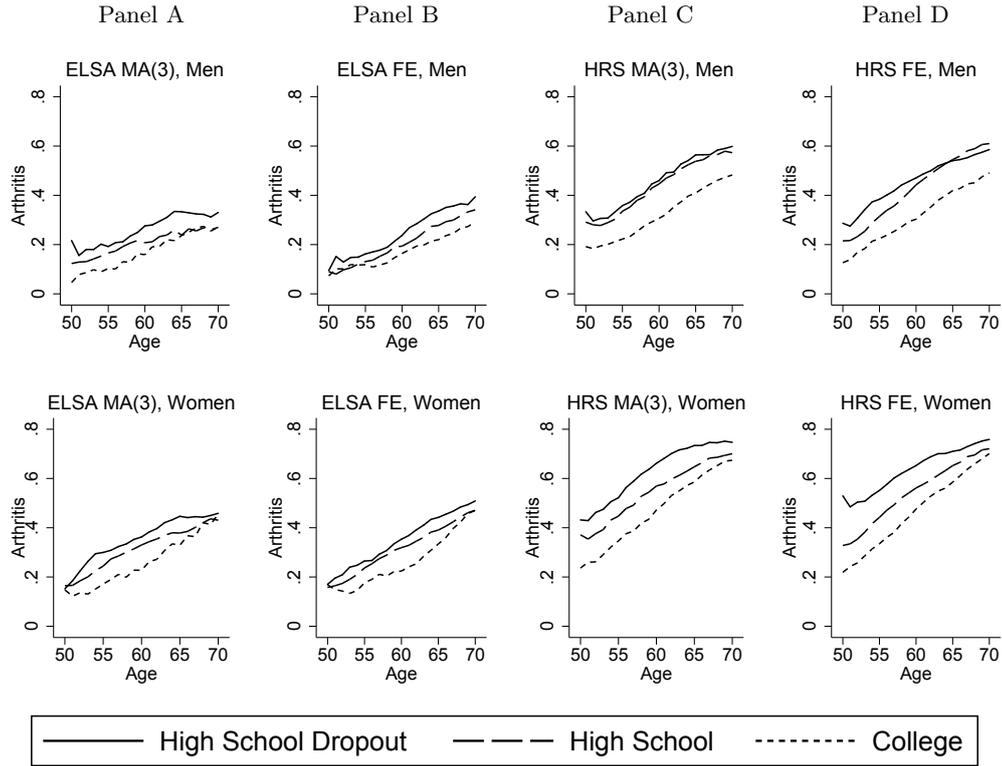


Figure 4: Prevalence of arthritis by age, gender and education. MA(3) indicates a 3-year moving average and FE indicates the exclusion fixed effects.

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 1.2 of the Online Appendix.

4.3 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 earlier findings in Banks et al. (2016) showing that Americans set lower thresholds for good and excellent health than do the English.

Curiously, and consistently with these differences in self-reported current health, 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.

Table 3: Subjective health variables, averages by gender

Variable	ELSA		HRS	
	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	44,478	58,741

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.

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.⁹ 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.

4.4 Cognition

High quality 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

⁹Plots for the each of the subjective measures can be found in Section 1.3 of the Online Appendix, while the weights assigned to each variable and the estimates from the first stage IV regression can all be found in Section 2.1 of the Online Appendix.

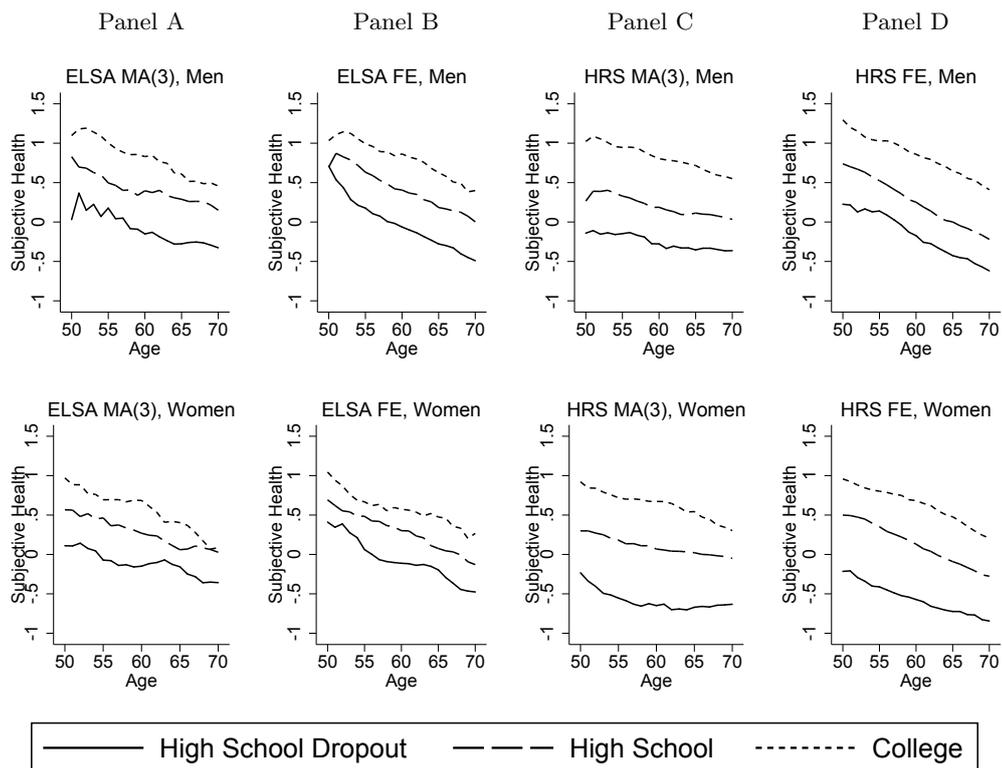


Figure 5: Single subjective health index by age, gender and education. MA(3) indicates a 3-year moving average and FE indicates the exclusion fixed effects.

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).¹⁰ Our focus is on fluid measures, primarily because they are available in both surveys across several waves,¹¹ 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 Amer-

¹⁰See 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.

¹¹ELSA does include a numeracy test in some waves (specifically, waves 1, 4 and 6), which might be considered a crystallized measure (and is used in Banks et al. (2010)).

icans 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.

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

Includes individuals aged 50-70.

Similar to the construction of our health index, we construct a cognition index that summarises the information content of the four cognition variables using Principal Component Analysis. The first principal component is plotted in Figure 6.¹² 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.¹³ 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

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

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

¹²Plots for each of the component variables are given in Section 1.6 of the Online Appendix, while the weights assigned to each variable can be found in Section 2.1 of the Online Appendix.

¹³We 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.

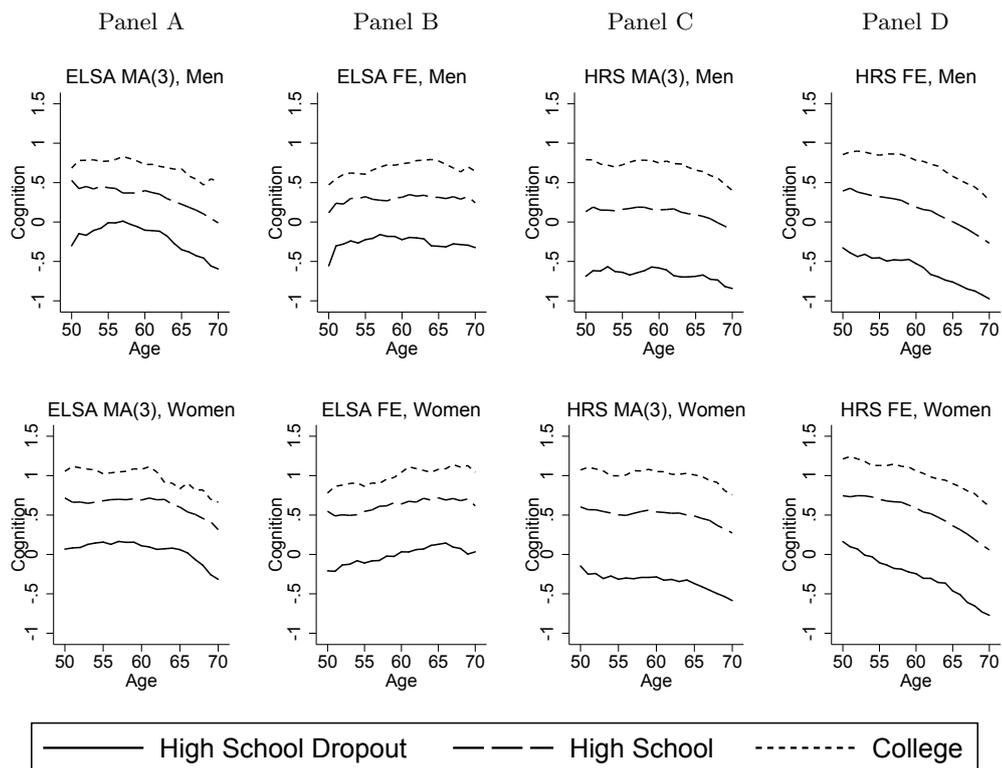


Figure 6: Cognition index by age, gender and education. MA(3) indicates a 3-year moving average and FE indicates the exclusion fixed effects.

their own or combined in an index, and we extend the model to include cognition. We also show the importance of allowing for initial conditions when estimating the impact of health.

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

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]
<i>Panel A: Employment on Subjective Health</i>								
High School Dropout	.196*** (.006)	.104*** (.007)	.207*** (.005)	.152*** (.006)	.132*** (.005)	.060*** (.005)	.170*** (.004)	.137*** (.005)
High School	.119*** (.006)	.057*** (.006)	.170*** (.003)	.119*** (.004)	.122*** (.005)	.070*** (.005)	.145*** (.002)	.112*** (.003)
College	.079*** (.008)	.053*** (.008)	.107*** (.005)	.080*** (.005)	.073*** (.009)	.048*** (.009)	.092*** (.005)	.078*** (.005)
<i>Panel B: Employment on Cognition</i>								
High School Dropout	.088*** (.006)	.017** (.007)	.085*** (.006)	.042*** (.007)	.056*** (.005)	.012** (.005)	.071*** (.004)	.037*** (.005)
High School	.033*** (.006)	.013** (.005)	.068*** (.003)	.032*** (.004)	.031*** (.005)	.006 (.005)	.062*** (.003)	.030*** (.003)
College	.013* (.007)	.002 (.008)	.050*** (.005)	.034*** (.005)	.018** (.009)	-.001 (.008)	.029*** (.005)	.017*** (.005)
<i>Panel C: Employment on Cognition and Subjective Health</i>								
<i>Cognition</i>								
High School Dropout	.037*** (.006)	.004 (.007)	.042*** (.006)	.026*** (.006)	.025*** (.005)	.005 (.005)	.031*** (.004)	.015*** (.005)
High School	.010* (.006)	.007 (.005)	.035*** (.003)	.019*** (.004)	.008* (.005)	-.001 (.005)	.030*** (.003)	.015*** (.003)
College	-.001 (.007)	-.004 (.008)	.032*** (.004)	.024*** (.005)	.005 (.008)	-.005 (.008)	.015*** (.005)	.008 (.005)
<i>Subjective Health</i>								
High School Dropout	.186*** (.006)	.103*** (.007)	.198*** (.005)	.149*** (.006)	.126*** (.005)	.060*** (.005)	.163*** (.004)	.135*** (.005)
High School	.117*** (.006)	.056*** (.006)	.163*** (.003)	.117*** (.004)	.121*** (.005)	.070*** (.005)	.139*** (.003)	.110*** (.003)
College	.079*** (.008)	.053*** (.008)	.101*** (.005)	.077*** (.005)	.072*** (.009)	.049*** (.009)	.089*** (.005)	.077*** (.005)
Sample sizes	4,692	4,692	5,777	5,777	6,957	6,957	9,199	9,199
	6,327	6,327	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 three variables summarised in Table 3 as well as the initial value of the health and cognition variables included in the regression. * indicates significant at 10%, ** 5%, *** 1%.

in the subjective health index is associated with 19.6% higher employment amongst high school dropout men; comparable estimates for high school graduates and college graduates are 11.9% and 7.9%, respectively.

However, estimates of the effects of subjective health 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 regression model, including health status during childhood, previous working experience, as well as health, employment status and cognition levels 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 8.8%, 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: Share of Employment Decline Explained by Cognition and Subjective Health

	Men		Women	
	ELSA	HRS	ELSA	HRS
<i>Panel A: Subjective Health</i>				
High School Dropout	.087*** (.015)	.136*** (.018)	.048*** (.008)	.109*** (.017)
High School	.048*** (.009)	.124*** (.008)	.052*** (.008)	.115*** (.006)
College	.045*** (.013)	.096*** (.013)	.024*** (.008)	.079*** (.011)
<i>Panel B: Cognition</i>				
High School Dropout	.002 (.003)	.036*** (.008)	-.003 (.002)	.057*** (.011)
High School	.001 (.002)	.030*** (.005)	0.000 (.001)	.033*** (.005)
College	0.000 (.002)	.037*** (.008)	0.000 (.002)	.017*** (.007)
<i>Panel C: Cognition and Subjective Health</i>				
High School Dropout	.087*** (.015)	.155*** (.019)	.046*** (.008)	.130*** (.019)
High School	.048*** (.009)	.139*** (.010)	.053*** (.008)	.130*** (.008)
College	.046*** (.013)	.118*** (.014)	.025*** (.008)	.086*** (.012)
Sample sizes	4,692	5,777	6,957	9,199
	6,327	18,756	7,911	29,905
	3,362	9,238	2,759	9,682

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

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 9). Estimates in Column 1 of Panel A show that the deterioration in health explains between 4.5% and 8.7% of the decline in men's employment in ELSA. The impact is larger 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, particular among those who leave education without qualifications.

Results for the HRS display similar patterns to those found in ELSA, only stronger (Columns

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 [1]	HRS [2]	ELSA [3]	HRS [4]	ELSA [5]	HRS [6]	ELSA [7]	HRS [8]
	<i>Panels A versus C of Table 6</i>							
High School Dropout	0.1	14.2	-3.4	18.9	0.476	0.001	0.228	0.007
High School	-0.3	12.2	0.3	12.5	0.443	0.000	0.435	0.000
College	2.8	22.7	2.8	9.0	0.349	0.000	0.363	0.097

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.

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 2 to 4 times larger than that of cognition decline (Panels A and B). Moreover, cognition explains about 2 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 indeed cognition increases modestly but significantly the explained share in employment decline in the HRS; in line with our earlier findings for ELSA, it 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 9). 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 also has little effect on the estimates.¹⁴ 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. This suggests that a single health index, if properly constructed, is sufficient for capturing the effect of health on employment.

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

	Men		Women	
	ELSA	HRS	ELSA	HRS
<i>Panel A: First principal component</i>				
High School Dropout	.087*** (.015)	.136*** (.018)	.048*** (.008)	.109*** (.017)
High School	.048*** (.009)	.124*** (.008)	.052*** (.008)	.115*** (.006)
College	.045*** (.013)	.096*** (.013)	.024*** (.008)	.079*** (.011)
<i>Panel B: Three subjective measures separately</i>				
High School Dropout	.074*** (.018)	.120*** (.020)	.033*** (.010)	.115*** (.018)
High School	.037*** (.011)	.128*** (.010)	.042*** (.009)	.131*** (.007)
College	.014 (.016)	.096*** (.014)	.031*** (.012)	.082*** (.012)
<i>Panel C: Health limits work</i>				
High School Dropout	.036** (.014)	.093*** (.019)	.015* (.008)	.109*** (.018)
High School	.024*** (.007)	.119*** (.009)	.022*** (.007)	.128*** (.007)
College	-.003 (.008)	.085*** (.012)	.004 (.007)	.078*** (.011)
Sample sizes	4,692	5,777	6,957	9,199
	6,327	18,756	7,911	29,905
	3,362	9,238	2,759	9,682

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

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,

¹⁴An 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.

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 [1]	HRS [2]	ELSA [3]	HRS [4]	ELSA [5]	HRS [6]	ELSA [7]	HRS [8]
<i>Panels A vs B (three separate subjective measures)</i>								
High School Dropout	-14.6	-11.6	-30.8	5.1	0.084	0.063	0.013	0.287
High School	-22.1	3.3	-20.7	13.5	0.081	0.196	0.025	0.000
College	-69.2	-0.3	26.4	4.9	0.006	0.482	0.234	0.250
<i>Panels A vs C (health limits work)</i>								
High School Dropout	-58.1	-31.4	-69.4	-0.4	0.000	0.002	0.000	0.485
High School	-50.4	-4.3	-58.4	10.7	0.000	0.188	0.000	0.009
College	-106.3	-11.6	-84.5	-1.0	0.000	0.101	0.004	0.461

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.

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 not systematically in the same direction. Accordingly, in most cases we fail to reject equality; in some cases we do reject, as for women with high school diploma in both ELSA and the HRS, but for modest relative differences. Rejection in these cases results from the contrasting parameters being strongly correlated – indeed they are overwhelmingly driven by the same information, which seems to be well summarised in the first principal component of health. Although rejecting the null in this context could be interpreted as evidence of non-linear effects of health, we tested for this possibility by contrasting our estimates from the linear employment model with those obtained from a probit specification and found no evidence in differences (see Table A10 in the Online Appendix).

However, we find that the information in a single observed measure significantly under-represents the variation in subjective health relevant for employment, particularly in ELSA. This is confirmed in the bottom panel of Table 9 for ‘Health Limits work’. 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.

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.

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

We deal with the potential justification bias and measurement error in the subjective health index by instrumenting it with the full set of objective measures. Objective measures may provide an incomplete picture of health status but they are likely to be strongly related to the subjective measures while being robust to justification bias. Moreover, measurement error and justification bias in subjective measures is likely to be unrelated to objective health. These features make the objective measures an ideal candidate for instrumenting the subjective health index. We first test their strength as 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 can 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/education/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 5.4.

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

	Men		Women	
	ELSA	HRS	ELSA	HRS
<i>Panel A: Subjective health</i>				
High School Dropout	.092*** (.021)	.148*** (.025)	.052*** (.013)	.142*** (.025)
High School	.056*** (.017)	.124*** (.013)	.057*** (.014)	.139*** (.010)
College	.057** (.024)	.149*** (.022)	.030* (.016)	.103*** (.019)
<i>Panel B: Subjective health and cognition</i>				
High School Dropout	.092*** (.021)	.161*** (.025)	.052*** (.013)	.151*** (.025)
High School	.055*** (.017)	.135*** (.013)	.058*** (.014)	.148*** (.011)
College	.060** (.025)	.162*** (.021)	.031* (.016)	.105*** (.018)

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

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, than similar IV estimates. The null hypothesis that the OLS and IV estimates are equal is not rejected at conventional levels in most cases. Where it is rejected, which only happens in the HRS, IV estimates are more noticeably larger than their OLS counterparts. We conclude 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 that contaminates the single index (despite it combining various measures of health).

Table 12 provides additional evidence on the validity of the single index assumption using the over-identification restrictions supplied by the many instruments we are using. If the objective

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 [1]	HRS [2]	ELSA [3]	HRS [4]	ELSA [5]	HRS [6]	ELSA [7]	HRS [8]
	<i>Panel A of Table 6 vs. Panel A of Table 10 (subjective health only)</i>							
High School Dropout	6.2	8.7	9.5	29.8	0.370	0.255	0.337	0.015
High School	16.4	0.4	8.5	20.7	0.284	0.482	0.333	0.002
College	27.2	54.8	23.4	30.8	0.257	0.001	0.334	0.058
	<i>Panel C of Table 6 vs. Panel B of Table 10 (subjective health and cognition)</i>							
High School Dropout	5.6	3.6	12.7	16.4	0.382	0.370	0.303	0.058
High School	15.7	-2.9	9.5	13.8	0.294	0.329	0.320	0.006
College	29.2	36.7	23.0	22.7	0.245	0.003	0.337	0.079

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.

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.¹⁵

Table 12: Overidentification Test

	Men		Women	
	ELSA [1]	HRS [2]	ELSA [3]	HRS [4]
<i>Panel A: Subjective Health</i>				
High School Dropout	0.165	0.248	0.191	0.000
High School	0.141	0.000	0.140	0.000
College	0.293	0.000	0.082	0.000
<i>Panel B: Subjective Health, with Cognition</i>				
High School Dropout	0.159	0.284	0.192	0.000
High School	0.150	0.000	0.144	0.000
College	0.293	0.000	0.074	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.

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 (6) and (7). The residuals were clustered at the individual level to account for serial correlation. In

¹⁵This is true for as long as the impact of health on employment is homogeneous, as explained in Imbens and Angrist (1994) and Angrist et al. (2000), but the exclusion restriction may not hold when a set of (individually valid) instruments is used if each identifies different margins of the effect. For this reason, the results from the Sargan test in Table 12 should be interpreted with caution.

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, although it is not rejected with the ELSA data.

Interestingly, adding the second principal component of subjective health to the employment regression does not noticeably change the predicted average effect of health (as shown in Table 8) but affects the test results. Results can be found in Table A11 of the Online Appendix; they show that we fail to reject the exclusion restriction for all but one group in ELSA once the second principal component is included in the model, but rejection remains high in the HRS. These results suggest that the impact of health on employment may indeed vary with health conditions, in line with the argument that it is the serious and persistent conditions that most affect employment. We further test whether this may be the case by restricting the objective instruments to the subset of most serious 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 A11 in the Online Appendix shows that the null is only rejected for three out of twelve cases (in each case, 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.

5.3 Assessing bias due to omitted objective health measures

We estimated the alternative model of health as a function of the entire set of objective measures in equation (3) 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 health 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 two groups, high school dropouts women in both ELSA and the HRS, the predictions from the full set of objective measures actually surpass those obtained using the subjective health index; however, the differences are small in both cases and only statistically significant for the HRS data.

Table 13: Share of Employment Decline Explained by Objective Health

	Men		Women	
	ELSA	HRS	ELSA	HRS
<i>Panel A: Blood pressure only</i>				
High School Dropout	.036** (.015)	.018*** (.006)	.012 (.010)	.035*** (.008)
High School	.011 (.010)	.019*** (.003)	.005 (.008)	.021*** (.003)
College	.006 (.009)	.008** (.004)	.008 (.010)	.010*** (.004)
<i>Panel B: Add Arthritis, Psychiatric, Lung</i>				
High School Dropout	.066*** (.018)	.105*** (.017)	.043*** (.013)	.128*** (.020)
High School	.023* (.013)	.065*** (.009)	.039*** (.011)	.078*** (.009)
College	.027* (.014)	.065*** (.014)	.009 (.016)	.056*** (.015)
<i>Panel C: Add Cancer, Diabetes, Stroke, Heart Attack</i>				
High School Dropout	.087*** (.021)	.152*** (.020)	.062*** (.015)	.179*** (.023)
High School	.033** (.015)	.087*** (.011)	.037*** (.012)	.099*** (.010)
College	.040** (.017)	.107*** (.018)	.016 (.017)	.076*** (.017)
<i>Panel D: Add Sight, Hearing – full specification</i>				
High School Dropout	.087*** (.021)	.147*** (.020)	.061*** (.015)	.190*** (.023)
High School	.033** (.015)	.090*** (.012)	.037*** (.012)	.102*** (.010)
College	.040** (.017)	.111*** (.019)	.022 (.019)	.076*** (.017)
Sample sizes	4,692 6,327 3,362	5,777 18,756 9,238	6,957 7,911 2,759	9,199 29,905 9,682

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

These results are consistent with the hypothesis that objective measures, even in rich datasets, provide an incomplete view of the health status affecting work capacity (recall discussion in Section 3.1). 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

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]
	<i>Panel A of Table 10 versus Panel D of Table 13</i>							
High School Dropout	-5.4	-0.5	17.3	34.0	0.406	0.488	0.234	0.015
High School	-40.1	-27.2	-35.4	-26.8	0.023	0.000	0.022	0.000
College	-31.0	-25.5	-28.6	-26.2	0.167	0.018	0.283	0.014

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.

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 mostly not statistically significant at conventional levels. These results align well with the findings in Bound (1991). 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 3 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).

5.4 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 (see Equation 9) 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 (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 .¹⁶

Table 15: Oaxaca Decomposition of US-English differences

	Men			Women		
	θ	ΔH	ΔY	θ	ΔH	ΔY
<i>Subjective health</i>						
High School Dropout	0.94	-0.01	0.08	0.84	-0.17	0.33
High School	0.98	-0.08	0.10	0.69	0.11	0.20
College	0.58	0.11	0.31	0.71	0.07	0.22
<i>Subjective health and cognition</i>						
High School Dropout	0.97	-0.03	0.06	0.84	-0.14	0.30
High School	0.92	-0.01	0.09	0.75	0.06	0.19
College	0.72	-0.01	0.29	0.81	-0.03	0.22

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.

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

¹⁶A description of the decomposition procedure can be found in Section 3 of the Online Appendix.

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.

6 Conclusion

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 well 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.

We do find that are estimates are sensitive to four important modeling 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 mortality 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 (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), pp. 644–650.

- 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. Müller, 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. Porapakarm (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), pp. 395–427.
- French, E. and J. Jones (2016). Health and retirement. 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 (2016). The effect of the affordable care act on the labor supply, savings, and social security of older americans.
- 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.
- 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.
- 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.
- McGarry, K. (2004). Health and retirement do changes in health affect retirement expectations? *Journal of Human Resources* 39(3), 624–648.
- 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.
- Pashchenko, S. and P. Porapakarm (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.
- 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.