I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11769

The Growing American Health Penalty: International Trends in the Employment of Older Workers with Poor Health

Ben Baumberg Geiger
René Böheim
Thomas Leoni

AUGUST 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11769

The Growing American Health Penalty: International Trends in the Employment of Older Workers with Poor Health

Ben Baumberg Geiger
University of Kent

René Böheim
WU Vienna University of Business and Economics, Austrian Institute of Economic Research, CESifo, JKU Linz and IZA

Thomas Leoni
Austrian Institute of Economic Research

AUGUST 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Growing American Health Penalty: International Trends in the Employment of Older Workers with Poor Health*

Many countries have reduced the generosity of disability benefits while making them more activating – yet few studies have examined how employment rates have subsequently changed. We present estimates of how the employment rates of older workers with poor health in 13 high-income countries changed between 2004-7 and 2012-15 using HRS/SHARE/ELSA data. We find that those in poor health in the USA have experienced a unique deterioration: they have not only seen a widening gap to the employment rates of those with good health, but their employment rates fell per se. We find only for Sweden (and possibly England) signs that the health employment gap shrank. We then examine possible explanations for the development in the USA: we find no evidence it links to labour market trends, but possible links to the USA's lack of disability benefit reform – which should be considered alongside the wider challenges of our findings for policymakers.

JEL Classification: J14, J18, H55

Keywords: disability benefits, employment of older workers, health employment gap

Corresponding author:
Ben Baumberg Geiger
School of Social Policy, Sociology and Social Research
University of Kent
Canterbury
Kent, CT2 7NF
United Kingdom
E-mail: b.b.geiger@kent.ac.uk

* Ben Baumberg Geiger was supported by UK Economic & Social Re-search Council (ESRC) [grant number ES/K009583/1]. Part of the research by René Böheim and Thomas Leoni was supported by the U.S. Social Security Administration through grant #1 DRC12000002-03 to the National Bureau of Economic Research as part of the SSA Dis-ability Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or the NBER.

1. Introduction

Since the 1980s, many high-income countries have seen a rise in disability benefit recipiency rates, often accompanied by a fall in the employment rates of older workers with ill-health (OECD, 2003, OECD, 2010, Whitehead, et al., 2009). In response to these trends, policymakers in many countries have reformed disability benefits programmes, reducing their coverage and generosity and making them more 'activating' (i.e. increasing their focus on return-to-work; see Figure 1 below). It has been assumed that these reforms will lead to improved employment rates for those with poor health relative to the rest of the population – yet almost no research has directly examined what has happened to these employment rates across time and place, and none have done this while taking steps to measure 'health' consistently over time.

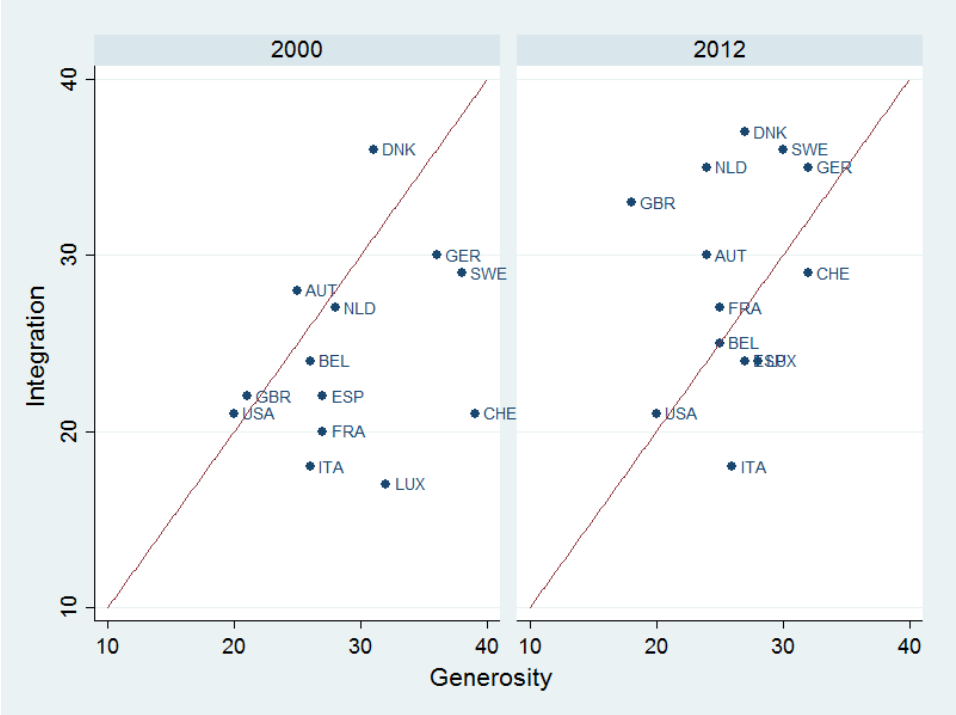


Figure 1: Trends in disability benefits policy in selected countries, 2000 to 2012.

Source: authors' calculations based on OECD (2003, 2010) and Böheim & Leoni (2018). Each scale is based on 10 sub-components ranging from 0 to 5, with a maximum of 50 points.

This is not an idle question, because it is unclear if the reforms will translate into improved relative employment rates for those in poor health, even if we ignore the partial overlap between poor health (as operationalised in research) and disability benefit receipt (Croda, et al., 2015). There is convincing evidence from natural experiments that reduced disability benefits eligibility leads to increased employment (e.g. de Jong, et al., 2011, Staubli, 2011). However, the overall long-term impact of these complex reforms is more ambiguous. For example, one form of activation is mandatory back-to-work planning (with the threat of sanctions), but the few existing randomised trials of mandatory interviews for disabled people show null or even negative effects on employment (Geiger, 2017). More broadly, van der Wel, et al. (2012) argue that benefit generosity is intrinsically linked within policy 'regimes' to social investment strategies that raise employment, and that it is therefore those countries with a combination of generous benefits and encompassing measures to support labour market (re-)integration that perform best.

Moreover, whatever the results of disability benefits reform in itself, trends in the relative employment of those in poor health may be outweighed by other factors, such as the business cycle (Benitez-Silva, et al., 2010) or other social security changes (Erosa, et al., 2012). Foremost amongst these are labour market factors; for example, part-time work (Jones, 2007, OECD, 2010:52) and job adjustments (Franche, et al., 2005) are crucial for the employment of those with poor health. It is difficult for older workers with health problems to find their way back into employment once they have lost their job, particularly once they receive a disability benefit (Burkhauser, et al., 2014, OECD, 2003:59). Hence the retention of existing workers may be a much more important influence on relative employment rates than the recruitment of disability benefit claimants.

In this paper, we examine employment trends of older workers in poor health across 13 high-income countries from the early 2000s to the mid 2010s. We show that most countries have seen stable relative employment of those with poor health, alongside improving abso-

lute employment levels. However, the USA is an exception: we find that the relative and absolute employment of those in poor health deteriorated in the USA. A careful analysis of the micro-data finds little sign that this is due to trends in part-time work, job tenure, or recruitment/retention, but that it may partly be due to the lack of disability benefits reform in the USA – notwithstanding some complexities in interpretation. Before this, however, we begin by reviewing existing studies and explaining why simple measures of health produce untrustworthy trends, before describing how we construct a more trustworthy measure (using the approach of Poterba, et al., 2013).

2. Methods and data

2.1 General approach

There is ample evidence that ill-health and disability reduce people's chances of getting and keeping work (Alavinia and Burdorf, 2008, Robroek, et al., 2013, e.g. Schuring, et al., 2007). A few within-country trends and cross-national comparisons do exist (Baumberg, et al., 2015, Geiger, et al., 2017), but these face a substantial methodological challenge, because different types of people have different ways of reporting their health across different times and places. For example, Jürges (2007) shows that differences in reporting style explain a large share of the differences in self-reported health across European countries. Indeed, health reporting is affected by both institutional characteristics (Angelini, et al., 2011) and whether respondents are working or not (Kalwij and Vermeulen, 2008 using SHARE data), making it difficult to interpret the few existing studies looking comparatively at health and employment.

A variety of strategies have been proposed to deal with these issues (e.g. Jones, 2006, Jürges, 2007). We use the approach suggested by Poterba, Venti, and Wise (2013) which is based on a series of specific health indicators that are combined into a single measure of latent health using principal component analysis (PCA). This index provides us with a fine-grained unidimensional measure of health. Poterba, Venti, and Wise (2013) document that

this health index is strongly related to mortality, as well as being a good predictor of future health events such as having a stroke. (In our analyses below, the health index is strongly correlated with self-reported work-limiting disability.) This index and related approaches have already been used in some influential comparisons, but the index has not been used for analysing health and employment over time. [1]

Our analysis depends on the assumption that each of the specific health indicators (listed below) is reported similarly over time within each country – it does not assume that these indicators are reported similarly across countries; nor does it assume that general measures of health are reported similarly within countries over time.

2.2 Data

To create the health index, we need high-quality comparative surveys of the general population that include a battery of specific health measures. Few such surveys exist, and like others (e.g. Poterba, et al., 2013), we therefore use three of the Global Ageing Datasets (a series of multipurpose panel surveys with extensive health and employment data, based on a common design): the USA Health and Retirement Study (HRS), the European Survey of Health, Aging and Retirement (SHARE), and the English Longitudinal Survey of Ageing (ELSA).[2] They provide uniquely detailed data on health and work that enable us to address our question, but they have the drawback that they constrain us to the 50+ population. The rest of this section focuses in turn on the sample, health variables, other variables, and analytical approach.

[1] Poterba et al use the index to underpin the latest phase of the National Bureau of Economic Research's 12-country comparison of 'Social Security Programs and Retirement around the World' (Wise, 2017) – which has been described as *"hands down the most influential use of international comparisons in economics"* (Banks and Smith, 2012). The approach is also similar to the Item Response Theory-based approach of the World Disability Report (World Bank and WHO, 2011).

[2] For a description of ELSA see Marmot et al. (2018); for SHARE see Börsch-Suppan (2018). The RAND HRS data were developed at RAND with funding from the National Institute on Aging and the Social Security Administration (RAND, 2018).

Sample

HRS, ELSA, and SHARE are longitudinal surveys of representative samples of several countries' populations aged 50+ (51+ in the case of HRS). HRS was the earliest of the surveys and, after annual collection 1992-1996, has been collected bi-annually since. The others began more recently and were based explicitly on the same model: ELSA has been collected bi-annually since 2002; and SHARE was first collected in 2004, and then approximately bi-annually since 2007 (with some variation by country). The surveys have been extensively used for comparative research (e.g. Avendano, et al., 2009, Cieza, et al., 2015, Crimmins, et al., 2010, Trevisan and Zantomio, 2016); we use the supplied survey weights, and full sampling details are given in the cohort profiles (Börsch-Supan, et al., 2013, Sonnega, et al., 2014, Steptoe, et al., 2013) and publicly accessible data documentation (Beaumaster, et al., 2017, Bugliari, et al., 2016, Phillips, et al., 2017).

To investigate trends, we use the earliest and latest available pairs of waves (we use pairs of waves in order to generate larger samples within the SHARE data), allowing us to construct trends from 2004-7 to 2012-15. Some countries are only intermittently included in SHARE over the study period, and we exclude countries for which the start/end sample sizes are too small to produce meaningful estimates; we also exclude Israel for which wider disability policy data are not available. Our resulting sample consists of Austria, Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Spain, Sweden, and Switzerland. For HRS and ELSA, we choose waves that overlap as closely with the SHARE waves; the resulting field and interview dates of the ELSA, HRS, and SHARE surveys overlap almost completely, as shown in Table 1 below.

	HRS	SHARE	ELSA
Baseline period	Wave 7: 2004/05	Wave 1: 2004/05	Wave 2: 2004/05
	Wave 8: 2006/07	Wave 2: 2006/07	
Latest period	Wave 11: 2012/13	Wave 5: 2013	Wave 6: 2012/13
	Wave 12: 2014/15	Wave 6: 2015	Wave 7: 2014/15

Table 1: Interview years of the included HRS, SHARE and ELSA waves.

Note: ELSA wave 3 (2006/7) is not used due to missing data for one of the health variables used below.

The Gateway to Global Aging Data (https://g2aging.org/, 'G2Ageing' data) provides harmonised versions of the three datasets. We added further data from the original HRS/SHARE/ELSA datasets to improve on the variables supplied in the harmonised dataset (e.g. employment and disability benefits receipt), and to add additional variables required for our analysis. Full details are given in Online Appendix A1, and are summarised below.

Health measures

Our latent measure of health is based on the following health indicators:

- 10 binary measures of motor skills (walking 100m/one block, lifting 5kg, pulling/pushing large objects, climbing one flight of stairs, climbing several flights of stairs, stooping/kneeling/crouching, picking up a small coin, sitting for 2hrs, getting up from a chair, reaching above shoulder height);

- 2 measures of functional disability, one for any limitation in Activities of Daily Living (ADLs), one for any limitation in Instrumental Activities of Daily Living (IADLs). The

former are basic and universal physical tasks such as eating, the latter are mixtures of physical and cognitive competencies such as preparing a hot meal (Breeze and Lang, 2006);

- 7 measures of self-reported doctor-diagnosed chronic diseases (high blood pressure, stroke, diabetes/high blood sugar, chronic lung disease (excluding asthma), heart problems, arthritis);

- 1 measure of fair/poor global self-reported health;

- 2 measures of non-optimal Body-Mass Index (BMI) – underweight and overweight – based on self-reported height and weight in HRS & SHARE and measured height and weight in ELSA;

- 2 measures of mental health, a scale measure treated as linear (CESD for ELSA/HRS, Euro-D for SHARE), and a binary measure of poor mental health based on the standard cut-off for the relevant scale measure.

Although the surveys are modelled on each other and the datasets are harmonized by G2Ageing, in some instances the concordance between the SHARE, HRS, and ELSA data is not perfect. Differences in terms of definition, construction or how the information was elicited sometimes required additional data harmonisation. In most instances, these adaptations were minor and we do not expect that they have further implications for our investigation. In some cases, however, the differences are more substantial, such as for the measure of mental health. We focus on within-country comparability over time to ensure that the measures are consistent within each country over time, but we do not assume that the measures are equivalent *between* countries.

Other self-reported measures

The validity of our analysis depends on having a consistent employment measure over time within each country. However, the employment variable in the G2Ageing harmonized data is not comparable within SHARE over time; we therefore created a revised, more consistent employment variable that is detailed in Online Appendix A1. In section 4 we also investigate self-reported disability benefit receipt; again, we improve on the G2Ageing version and create a more consistent, more precisely operationalised variable, in a series of steps that are detailed in Online Appendix A1. Alongside this, we also use the G2Ageing variables on work hours, tenure, age, and gender.

Institutional measures

Separately to our micro-level data, we consider aggregate-level trends in disability benefits policy using the influential OECD policy scales, as shown in Figure 1 above. The OECD use two policy indicators, each of them consisting of ten sub-dimensions measured according to a predefined scale which ranges from zero to five points (OECD, 2010):

1. *'Generosity' (aka 'Compensation')*, for which higher scores indicate greater generosity (including the coverage and level of disability benefits, the minimum degree of incapacity needed for benefit and full benefit entitlement, the type of medical and vocational assessment, as well as information on sickness benefits);
2. *'Activation' (aka 'Integration'),* for which higher scores mean a more active and employment-oriented approach (including the complexity and consistency of benefits and support systems, the degree of employer obligations towards their employees, the timing and extent of vocational rehabilitation, and the existence of work incentives for beneficiaries).

The OECD scores are only available for the period 1990 to 2007. We therefore use updated scores provided by Scharle et al (2015) and Böheim and Leoni (2018) that take into account changes up to the year 2013.

2.3 Analytical approach

International comparisons of older workers' employment are complicated by differences in retirement ages. One option is to restrict the analysis to a specific age group (e.g. 50-59 year olds), but this would exclude an important segment of the workforce in countries with a high statutory retirement age. Furthermore, this would not necessarily ensure comparability, because the labour market situation of persons of the same age in different countries could still be different depending on the time until statutory retirement. Since we are interested in how far different countries integrate the least healthy parts of their workforce, we instead focus on individuals who were between 50 and the respective statutory/regular pension age in the country at the time of the survey.[3]

This leaves us with a sample of 148,293 observations for 65,171 individuals (of which 110,345 observations are in the baseline/latest periods; see Online Appendix A3 for sample size per country/wave). While this is a large sample overall, the sample sizes for the SHARE countries are often small (particularly when we restrict our analysis to single tertiles of the health distribution), and we are therefore unable to stratify our analyses by gender. Incomplete data is generally low for all survey-waves with the exception of HRS (particularly the earlier waves, primarily for mental health and certain ADL/mobility limitations) and ELSA (for BMI, which required physical measurements and was only undertaken at alternate waves), as shown in Online Appendix A2. We consider the role of missing data in two sensitivity analyses, one of which excludes BMI to reduce missingness in ELSA, the other of which

[3] For the US, we imputed information on the normal retirement age, differentiated by year of birth, as indicated by the U.S. Social Security Administration. For England, state pension ages by year of birth were taken from Government sources. For SHARE countries, retirement age is included at the individual level for those countries/respondents within the SHARE Job Panel dataset. We updated this information for all countries/respondents in our sample using the documentation provided by the OECD (Pensions at a Glance), the EU (Ageing Reports) and the MISSOC database; depending on the country, the statutory retirement age of an individual might depend on demographic characteristics such as gender and year of birth. Descriptive statistics on retirement ages by country-wave are given in Online Appendix A3.

performs a full multiple imputation analysis; this is described below and in Online Appendix A4.

The health index is generated using a PCA of the health variables listed above. We use the first principal component, which represents the weighted average of the health indicators (where the weights are chosen to maximize the proportion of the variance of the individual health indicators that can be explained by the first principal component[4]). The first principal component can be interpreted as a latent health index. The PCA is carried out separately for each country, but we pool all waves and use the population aged 50 to 67 years to maximise sample size and to ensure that each health indicator is consistently weighted over time. We use this to construct percentile scores of each individual's position within the health distribution (within that country-wave); for most analyses, we group the scores and consider those in the bottom tertile as being in 'poor health' in that country. This creates equally-sized groups of those with *relative* poor health in each wave, rather than differently-sized groups with similar *absolute* health.

To examine changes over time net of demographic change, we estimate the impact of health on employment status, y_{ijk}:

$$y_{ijt} = \boldsymbol{\beta}_1 health_{ijt} + \boldsymbol{\beta}_2 country_j + \boldsymbol{\beta}_3 (health_{ijt} * country_j) + \boldsymbol{\beta}_4 t + \boldsymbol{\beta}_5 (health_{ijt} * t) + \boldsymbol{\beta}_6 (country_{ijt} * t) + \boldsymbol{\beta}_7 (health_{ijt} * country_j * t) + \boldsymbol{\beta}_8 (age_{ijt} * country_j) + \boldsymbol{\beta}_9 (age_{ijt}^2 * country_j) + \boldsymbol{\beta}_{10} (gender_{ijt} * country_j) + \boldsymbol{\beta}_{11} + \epsilon_{ijt}$$

[4] Using the STATA 'pca' routine, the eigenvectors are returned in orthonormal form, that is, uncorrelated and normalized.

where health indicates person i's rank in country j's health distribution at time t.[5] For the main set of estimates, we distinguish between the tertiles of the health distribution (below, we also use a finer distinction and use deciles). While our data are clustered, we are not interested in partitioning the variance between the individual and the country-wave level, and the low number of countries in our sample would likely lead to biased estimates in a multilevel analysis (Bryan and Jenkins, 2016); we therefore account for clustering using cluster-robust standard errors. We use these estimates to estimate the marginal effects of health which form the basis for our discussion below.

[5] We pool observations over time rather than to estimate fixed-effects panel estimates. In our main analyses we use OLS which should produce effectively identical results to logit models (Hellevik, 2009) but are more easily interpretable and better-validated for certain multiple imputation analyses; however in sensitivity analyses we use logit models that produce effectively identical results (see Online Appendix 3).

3. Trends in the employment of persons with poor health

In all countries considered here, health is an important determinant of employment: employment rates for those with poor health are considerably lower than those with better health (see Online Appendix 4). We estimate the gap in employment rates using the model displayed above, controlling for the age-gender structure in each country and period. The estimated gap between the bottom and top health tertiles ranges from 10.9% (percentage points) in Switzerland 2012-15 to 41.9% in the USA 2012-5. Our main focus here however is on changes over time. The trend in the employment gap from the start (2004-7) to the end (2012-15) of this period is shown in Figure 2 below. (The accompanying Table is given in Online Appendix A4.)

For most countries, Figure 2 shows no evidence of a systematic change despite widespread disability benefit policy reforms. However, there is some evidence of a systematic trend in three countries. The evidence suggests that the gap decreased by 7.4% (95% confidence interval (CIs) -13.6 to -1.3%) in Sweden. In contrast, there is evidence that there were sharp deteriorations in the position of those with poor health in Austria (by 10.3%, 95% CI 0.8 to 19.4%) and in the USA (where the gap rose by 8.1%; 95% CI 4.8 to 11.4%).

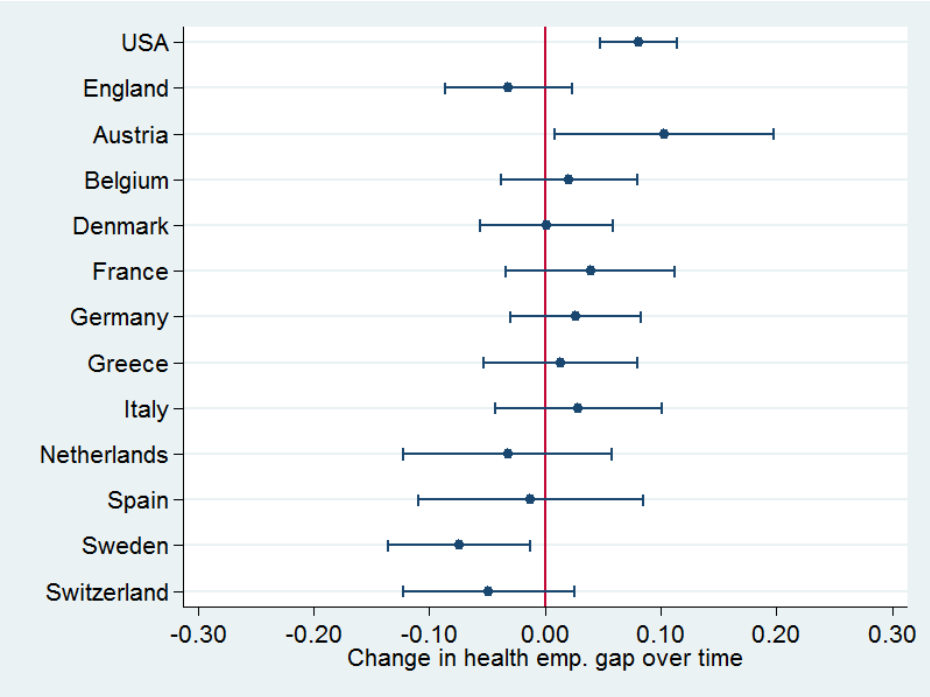


Figure 2: Gap in employment rates between bottom vs. top health tertile – change from 2004-7 to 2012-15

Note: Rising health employment gap = an increase in the difference in employment rates between those in poor health vs. those in good health. Gaps are estimated controlling for age and gender.

These comparisons indicate how those in poor health fared relative to those with better health. To see how they fared *per se,* we look at *absolute* changes in the employment rate of those in poor health, shown in Figure 3 below. This shows a generally positive development in most countries. We observe an increase in employment rates (after controlling for demographic changes), although in some countries these shifts are not statistically significant at conventional levels (and Greece experienced a slightly negative but not statistically significant change). The greatest increase in the employment share of workers with ill health took place in Sweden, followed by Switzerland, the Netherlands, Belgium, and Germany. The general increase may well be the consequence of increasing labour force participation of women, as well as labour market or pensions policy changes.

The US, however, is an outlier. The development in US was unique as it was the only country in our sample where both the relative *and* the absolute employment of older workers with ill health declined. Although in Austria the employment gap rose (Figure 2 above), the employment of those with poor health increased (Figure 3 below); the rising gap is because this increase in employment was much smaller than for those in better health.

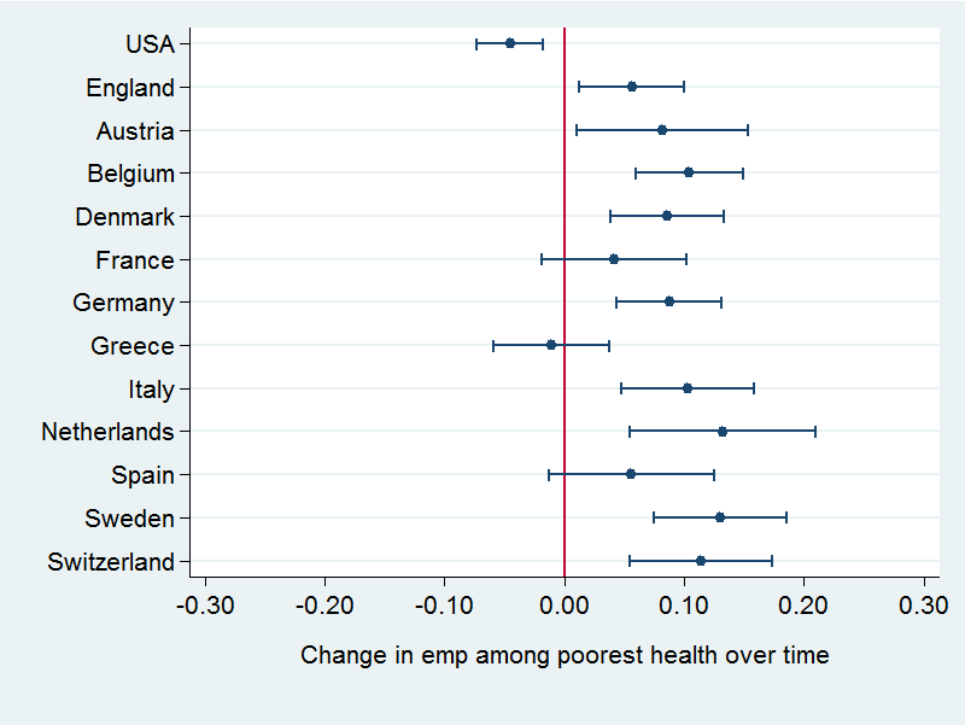


Figure 3: Change in absolute employment level of respondents, in bottom health tertile, 2004-7 to 2012-15.

Note: A positive value indicates higher employment rates over time for persons who were in the bottom third of the health distribution. Employment levels are estimated controlling for age and gender.

3.1 Sensitivity analyses

Before we explore the anomalous position of the USA further, we first demonstrate that these results are robust through a series of sensitivity analyses. (The full results are in Online Appendix 4.) Firstly, we re-estimated the analyses with logit rather than OLS models, and (unsurprisingly) the results are effectively identical. Secondly, because refreshment samples are not added to the surveys at every wave, the minimum age at different waves changes (as shown in the changing mean ages in Online Appendix A3). We therefore conduct a further analysis on a sample of those aged 54 (not 50/51) to retirement age. Again, our results are robust: the USA is unique in being the only country in which there both the absolute and relative employment situation of those in poor health deteriorated. Results for some other countries change, however, including Austria where the increasing health employment gap is more marked (14.5% rather than 10.3%).

Our remaining sensitivity analyses deal with missing data. The greatest source for missing data is BMI, particularly in England; we therefore re-run the analyses excluding BMI from our

latent health measure. As an alternative approach, we maintained our use of BMI but used a multiple imputation analysis to account for missingness in all variables (for further details see Online Appendix A4). These analyses had most impact on the results for England: both sensitivity analyses suggested more positive trends, with the absolute employment trend among those in poor health rising by 7.6-7.7% rather than 5.6%, and also providing some suggestions that the health employment gap declined (by 4.4% (-8.9% to 0.04%) in the multiple imputation analysis and 5.2% (-10.3% to -0.1%) when excluding BMI, compared to 3.2% (-8.6 to 2.3%) in the main analyses above). For Austria, the sensitivity analyses confirm the main results about the changes in employment levels and gaps, but some sensitivity analyses led to smaller and less precise coefficients for the health-related employment gap. At the same time, these sensitivity analyses resulted in *larger* and more precise coefficients for the absolute trend in employment among those in poor health. Otherwise, however, the substantive conclusions of both sensitivity analyses are identical to our main analyses.

3.2 Distribution of employment along the whole health distribution

One advantage of having a fine-grained health index is that we can examine trends in employment across the full distribution of health. Rather than comparing employment rates by tertile of health, we treat health as continuous (using 10 deciles) and allow a flexible (cubic) specification for the relationship between health and employment in each country-wave, controlling for age and gender. Results for the countries with changing relative employment rates are shown in Figure 4. (Remaining countries are shown in Online Appendix A4.)

The analysis of the changes across the full health distribution reveals two main findings. Firstly, it demonstrates once more that those with poor health have lower employment rates than those with better health. However, we see that this effect is often concentrated in the bottom half of the health distribution, such as in the US, Sweden and England. In some countries (e.g. England, and Switzerland in Online Appendix A4), the extent of this is such that the employment shares are low for those at the bottom of the health distribution, but once people have a certain level of health, their chances of being employed varies little with

improved health. In Austria, on the other hand, each change in health has a similar link to employment across the health distribution.

Secondly, Figure 4 provides more detail about the deteriorating employment position of those with poor health in the US. Along the entire health distribution, we do not find any improvement in the chances of being employed – but there was a sharper fall in employment rates for those in the bottom half of the health distribution. The increasing health-related employment gap in Austria is different; here there were increases in employment rates along the whole health distribution, but only minor improvements for those with the poorest health. And in Sweden (and to a lesser extent in Germany and Switzerland; see Online Appendix A4), the employment shares of those in the bottom half increased more-than-proportionally over time. Between these extremes we see countries which changed little such as England who has the most stable relationship between health and employment across waves.

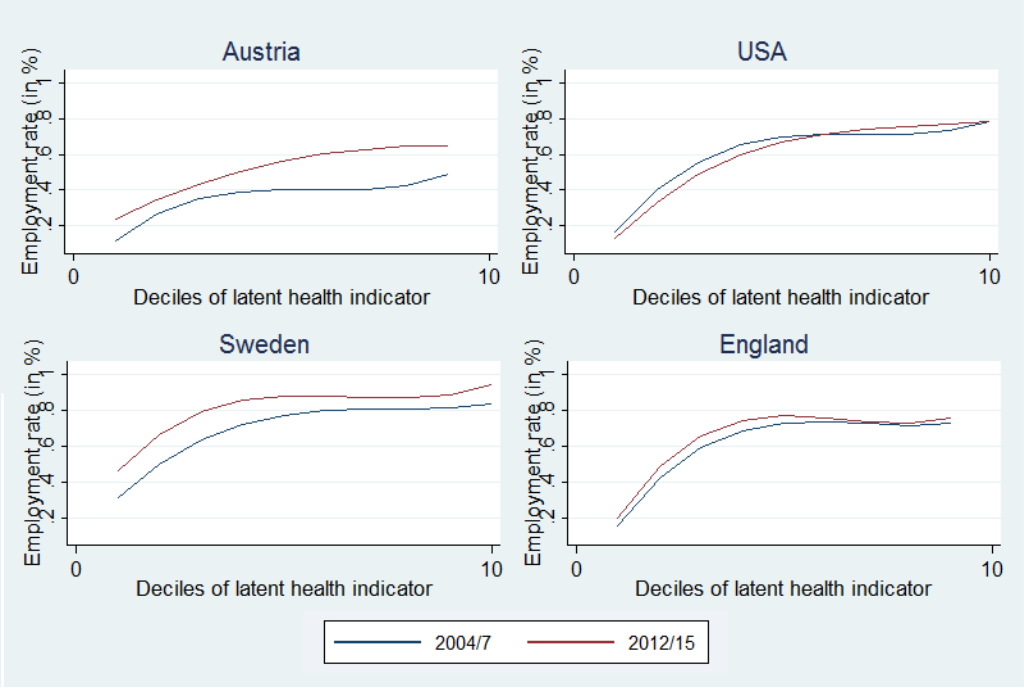


Figure 4: Employment across the whole distribution of health, 2004-7 to 2012-15.

Note: Figures show the association between employment and a cubic function of health, controlling for age and gender.

4. Exploring trends: disability benefit policies

We have seen that in many countries, employment among older working-age people has increased. Yet the results so far contradict the expectation that disability benefit reform would reduce the health employment gap – reform has taken place, but the gap does not seem to have reduced in most countries (Sweden being the solitary counter-example). In Austria and the USA, the situation even deteriorated. In this section we try to understand this by examining trends in disability benefit receipt and the association with changes in employment status using the rich micro-data of the ageing surveys.[6] For ease of presentation, we focus on a geographically and institutionally dispersed subsample of countries that show different trends (Austria, Sweden, Denmark, England, Germany, Italy, and the US).

[6] For this section only we use logit rather OLS models, because to the extent that OLS and logit models diverge, this will be particularly apparent for rare outcomes such as disability benefit receipt.

4.1 Overall disability benefit receipt

Figure 5 shows that there are substantial differences in the development of disability benefit receipt across countries. Sweden – and to a lesser extent Denmark, Italy, Switzerland, and England – experienced a reduction in disability benefit claim rates among older working-age adults. Conversely, we see that in Austria, Germany, and the USA receipt increased over the period.

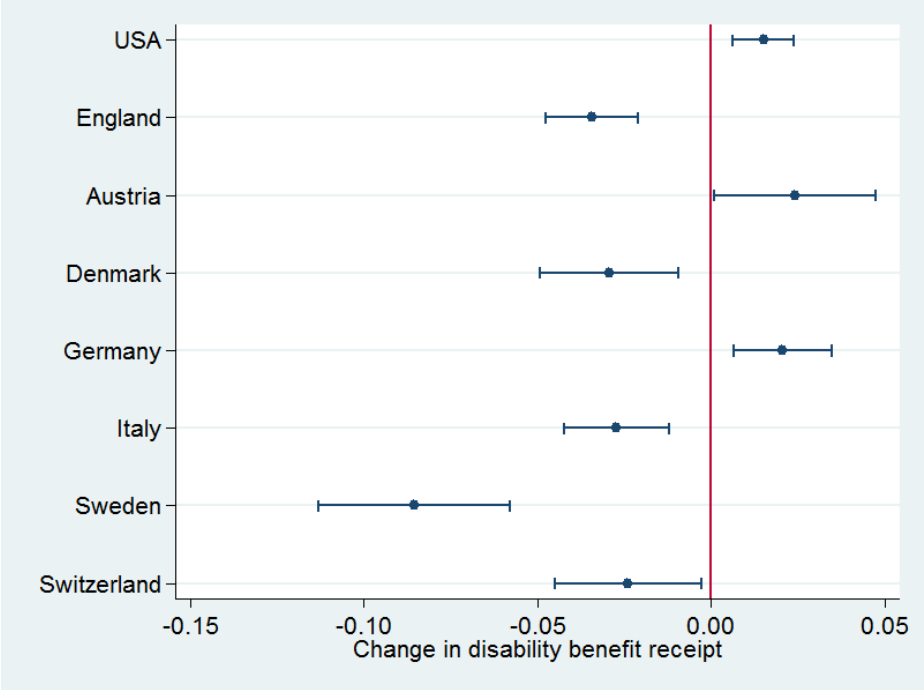


Figure 5: Disability benefit claim rates, 2004-7 to 2012-15.

Note: A positive value indicates higher disability benefit claim rates over time across the whole population, controlling for age and gender.

There is some overlap between the trends in employment rates and the trends in disability benefit claim rates among those with poor health: both claims and the health-related employment gap fall in Sweden, and both of these rise in the USA (and more tentatively in Austria). However, this does not necessarily indicate a causal link from disability benefit claims to non-employment. Moreover, there is not necessarily relationship between these trends. The health-related employment gaps for Germany, Italy, Denmark, Switzerland, and England have been effectively static, despite non-negligible changes – in contrasting directions – in disability benefit receipt.

4.2 Further complexities around disability benefit receipt, work and health

One caveat on Figure 5 is that it considers the whole population, rather than just those in poor health – and we have already noted that disability benefits are targeted at those in poor health to differing extents in different countries (Croda, et al., 2015), notwithstanding the incomplete observability of health/disability in the micro-data. In Online Appendix A5, we disaggregate the trend in disability benefit receipt by health. This emphasises that disability benefit trends are focussed on those in poor health, but we nevertheless do see small significant trends in disability benefit receipt among the top two health tertiles in Denmark, Italy, Sweden (all showing declines), and Austria (showing a rise).

A further finding is that in some countries, there are non-negligible proportions of disability benefit claimants who also work (even if this is prohibited in other countries) – this has been briefly noted elsewhere (Hogelund, 2003:161, OECD, 2003 Chart 3.7), but nevertheless seems to be ignored in wider policy debates. There is therefore a further possible break in the disability benefit-employment link: if people who receive disability benefits are already working, then any changes in their disability benefit receipt cannot make them more likely to be in employment. We therefore split trends in disability benefit claim rates by claimant working status and show the results in Online Appendix A5. The trends for simultaneously claiming and working are imprecisely estimated given the small numbers involved in many countries, but it seems that some reforms target working claimants differently to non-working claimants. For example, Denmark reduced the number of non-working claimants while the number of employed claimants did not change, whereas Sweden seems to have reduced the numbers of both types of claimants.

Policy, disability benefit receipt, and employment

We should finally analyse the role of disability benefits *policy,* rather than disability benefits *receipt,* on the employment of persons with poor health. As we showed in Figure 1, many countries have not only been reducing the generosity of their benefits, but have also been trying to make them more activating. It is perhaps unsurprising that the majority of countries have therefore seen a decline in disability benefit claims. Yet the USA and Austria stand out

once more here: the only country that has implemented weaker reforms than Austria is the USA, whose policy in this area has been effectively static (for one discussion of the politics of this, see Morris, 2016).

We should stress that these observations are not meant to substitute for more detailed policy evaluations (such as those that we have cited above), but rather aim to complement such evaluations by asking different questions of different data. We return to the relationship between these different approaches in the Discussion below.

5. Exploring trends: labour market factors

A further explanation of different trends in different countries comes from labour market factors: as we argued above, it is possible that any effects of disability benefits policies are outweighed by wider changes. We explore two factors that are likely to be important: hours of work, and transitions into/out of work.

5.1 Hours of work

While part-time work is more common among workers with poor health (see Online Appendix A6), there is little evidence that changes between part-time and full-time work explain different trends across countries. Figure 6 displays trends in the employment of workers with ill health, split by the number of regular working hours per week. (The table is given in Online Appendix A6.) The number of workers who work less than 15 hours/week did not change in most countries in our sample, with the exception of Germany (which saw greater rises in small part-time than in larger part-time or full-time work). We do observe increases in the number of workers who worked between 15 and 29 hours/week in some countries, and a contrasting trend in Sweden. However, we generally find that trends in employment are similar to trends in full-time work (30+ hours/week): they generally rose (particularly in Sweden), but fell in the USA.

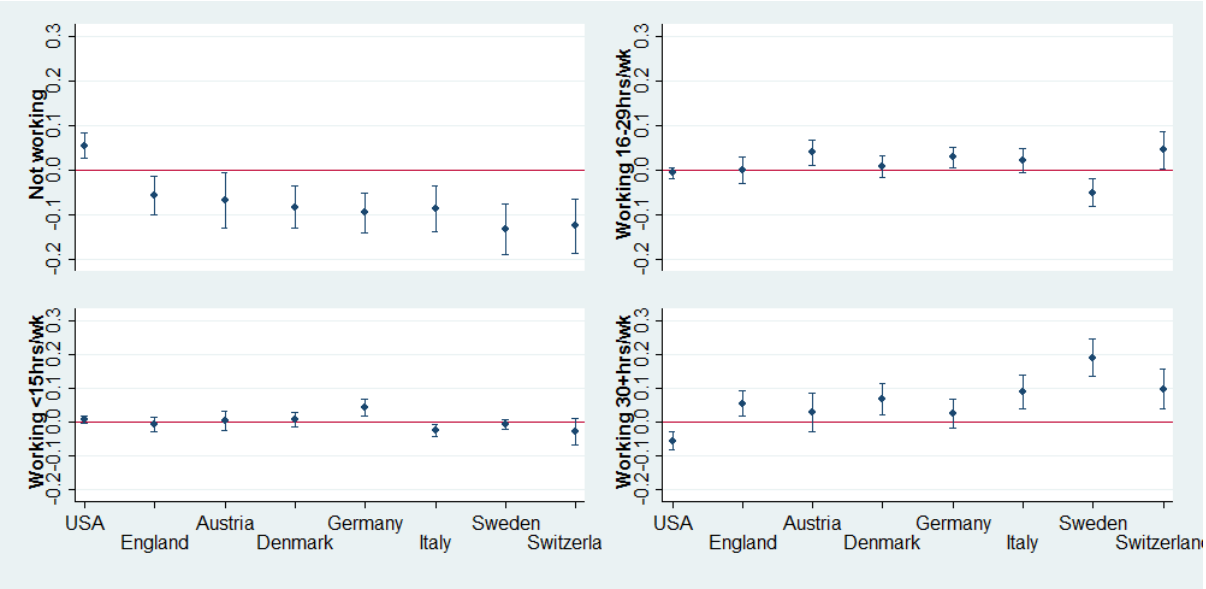


Figure 6: Change in work status, by hours worked 2004-5 to 2014-15, bottom health tertile.

Note: A positive value indicates rises in the particular working category (labelled on the y-axis), controlling for age and gender.

5.2 Transitions in/out of work

We previously suggested that the retention of workers with poor health might be much more important for relative employment rates than the recruitment of disability benefit claimants. There is also reason to believe that these will differ internationally, given evidence that there are national cultures of human resource strategies and indeed of 'employment regimes' more broadly (Gallie, 2009).[7] In this section, we therefore analyse if there is any evidence for country-specific changes in recruitment/retention patterns (potentially indicative of changing human resource management cultures) that may partly explain the USA's unique trajectory.

One way of examining this is through job tenure – the length of time that current workers have been in their present job – to see whether those who have poor health have different

[7] For example, employers in Denmark, Sweden, and the Netherlands have a greater propensity to use accommodation and development strategies than those in Germany, Italy, and Poland (Van Dalen, et al., 2014). Danish employers are particularly unlikely to use measures that favour labour market exit and, overall, situations in which firms experience recruitment problems were conducive to the implementation of accommodation and development strategies for the retention of older workers. Results for the Netherlands indicate also that – particularly in times of crisis – firms reduce the recruitment of older workers significantly, but that they also tend to make efforts to retain older workers and "spare" them from layoffs.

job retention probabilities. (The table is given in Online Appendix A6.) We find only modest differences in tenure between persons with different health levels in some countries, and no statistically significant differences in others (such as Belgium and France). Much more pronounced, however, are differences across countries: the Anglo-Saxon countries have the shortest job tenures (an average of around 10 years per older worker), noticeably lower than the Scandinavian countries (14 and 17 years in Denmark and Sweden respectively), which are still below those in Central Europe and Italy (ranging from 17 to 24 years). Yet while this is crucially important for understanding cross-national differences at any one time, it does not seem to explain divergent trends: there is little change over time in either overall or health-stratified job tenure, as shown in Figure 7 below.

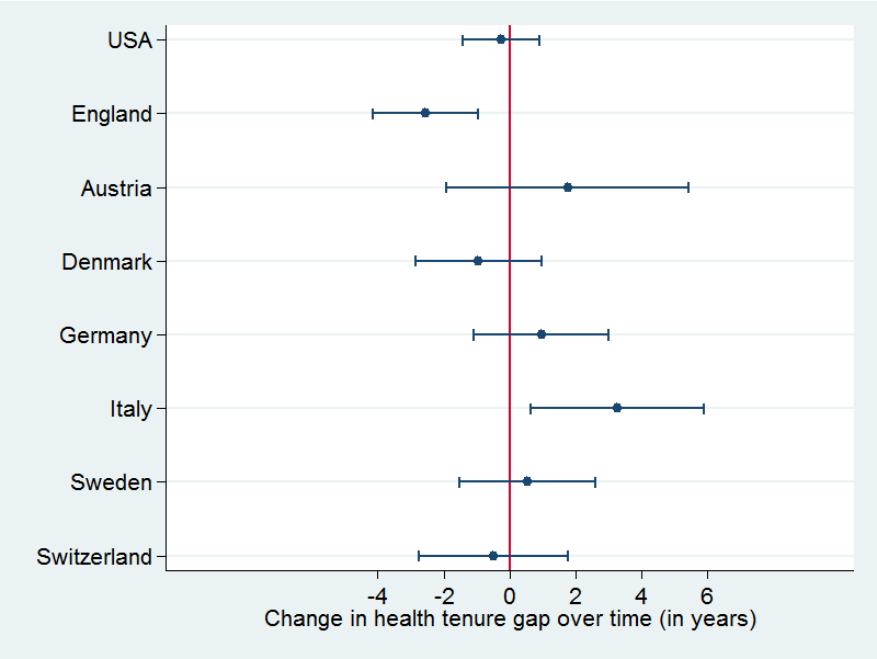


Figure 7: Change 2004-5 to 2014-15 in the tenure gap between the bottom health tertile and other health tertiles.

Note: A positive value indicates rises in tenure among those in poor health, controlling for age and gender.

An alternative way of investigating this is to make use of the longitudinal structure of the data and examine labour market transitions between waves; we focus on changes between the first pair of waves (2002-3 to 2004-5) and the last pair of waves (2012-13 to 2014-15). If we first look at the recruitment rate (the left-most columns of Table 2), we see that there have been relatively few changes in the recruitment rate over time (there has been a sizeable drop in Switzerland, where at the same time however baseline employment increased from 69.9% to 77.3%). What is most striking about the retention rate is that it is much greater than the recruitment rate – across countries people are consistently more likely to stay working than they are to move from non-work into work. There are also several countries that have seen substantial increases in the retention rate over time, namely Austria, Denmark, Germany, Sweden, and Italy.

Again, though, these do not provide clear explanations for the wider trends in employment rates of those in poor health. We have seen that there were increases in absolute employ-

ment rates of those in poor health over time, but this coexists with wide variation of recruitment and retention trends. Moreover, the clearest difference between the USA and other countries is in trends in the baseline employment rate. For most countries, the reason that employment rates increased at the follow-up wave (i.e. that more people were working in 2014/15 than 2006/7) is that employment was higher at the baseline wave (more people were working in 2012/13 than in 2004/5), rather than because the wave-to-wave transition rates changed noticeably. In the USA, in contrast, fewer people were working in 2012/13 than in 2004/5 – and it is this (rather than the slight declines in the recruitment and retention rates) that drives their unique deterioration.

	Recruitment rate[1]		***Retention rate[2]***		***Baseline[3]***	
Country	*2004/5➔ 2006/7*	*2012/3➔ 2014/5*	*2004/5➔ 2006/7*	*2012/3➔ 2014/5*	*2004/5➔ 2006/7*	*2012/3➔ 2014/5*
USA	7.0	6.1	80.1	78.9	47.9	38.3
England	5.2	5.5	81.2	81.4	50.8	52.6
Austria	4.8	4.5	61.8	72.3	38.4	41.9
Denmark	8.9	12.0	77.5	84.1	52.0	56.8
Germany	12.5	14.0	70.6	81.3	47.1	54.6
Italy	10.4	7.1	69.1	81.7	36.7	43.5
Sweden	11.0	12.7	80.9	90.3	61.7	69.4
Switzerland	31.3	23.7	89.2	88.6	69.9	77.3

Table 2: Changing employment transitions between waves for those in poor health.

Note: Numbers are percentages. [1] 'Recruitment rate' refers to the proportion of those out-of-work at the baseline wave who move into work at the follow-up wave; [2] 'Retention rate' refers to the proportion of those working at the baseline wave who stay working at the follow-up wave; [3] 'Baseline' refers to the proportion of respondents who are working at the baseline wave.

Finally, we can further examine Table 2 by separating out different non-working statuses, which is shown in Online Appendix A6. Figure A5 shows that in a few countries (notably Switzerland, Denmark, and Germany) the share of persons who started to work between

waves increased, however these changes were modest. We also observe a decrease in the share of retirees among the non-working population in some countries. The shift is particularly pronounced in Austria and can be linked to the pension reforms that were implemented in 2000 and especially 2004 with the aim of restricting access to early retirement (Busemeyer, 2005). The strong increase in unemployment suggests that, for a considerable number of those affected by the pension reforms, a substitution between early retirement and unemployment took place.

6. Conclusions

Employment trends among older workers with poor health

Rising disability benefit recipiency and falling employment rates (particularly among older workers with ill-health) since the 1980s prompted many countries to reform their disability benefit systems. These reforms generally reduced the generosity of benefits while making them more activating. (See Figure 1 above.) Yet few studies have examined how employment rates of older workers with poor health have changed across countries. We present new estimates of changing relative and absolute employment rates of older workers (aged 50/51 to variable retirement ages) in 13 high-income countries from 2004-7 to 2012-15 using ELSA/SHARE/HRS data. In order to obtain a consistent measure of health, we construct a latent health index from multiple specific health indicators using the method suggested by Poterba, Venti and Wise (2013).

We find that the USA is exceptional among the countries for which we have comparable data: it has seen a unique deterioration in the employment position of older workers with ill-health, compared to the other 12 countries we consider here. Not only did the gap between the employment rates of those with poor health and good health increase in the USA (by 8.1%; 95% CI 4.8-11.4%), but the employment rates of those with poor health *per se* fell (by 4.5%, 95% CI 1.8-7.3%). In only one other country, Austria, did the health employment gap increase (although this was not significant in all specifications), and in no other country did

the absolute employment rates of older workers with poor health fall. Indeed, in most countries (with the exception of Greece) employment rates rose – even if these rising employment rates were often similar to those among older workers in better health. Only in Sweden (and possibly also England, as we find such evidence in some but not all specifications) are there signs that the health employment gap fell over this period.

We note several limitations of our analysis. While the Poterba, Venti, and Wise (2013) method for measuring ill-health does not require the assumption that survey respondents interpret their general self-reported health consistently across time, it nevertheless requires that they consistently interpret that more specific health indicators over time. Although this assumption is less strict and permits a substantial improvement on using a single general health indicator, it should be borne in mind. However, we show that our results are generally robust to a variety of sensitivity analyses (including a multiple imputation analysis to deal with missing data).

The available data restrict our sample to older people; while ill-health is concentrated among older workers, this may nevertheless conceal different trends among younger people with ill-health. While this might cast doubt on the external validity of our results, there is evidence that the disability employment gap in the US increased for the entire 16-64 age group, and that people with disabilities were considerably more likely than others to experience involuntary job loss 2007-13 (Mitra and Kruse, 2016). [8]

[8] Note that the trend is non-linear: the disability employment gap rose from late 2008 to 2011 and stayed elevated until 2016, but it has since declined slightly, and the two earliest periods in the data (Jun-July 2008) show a higher disability employment gap. This 2008-2018 data are from the historical tables of labor force statistics Table A-6 from the US Bureau of Labor Statistics, https://www.bls.gov/webapps/legacy/cpsatab6.htm [accessed 5/6/2018]; the standard US disability measure is based on any self-reported activity limitation/serious hearing difficulty/serious difficulty seeing. There is also a slight increase in the disability employment gap 2006-2010 in Livermore et al (2015).

Exploring these trends

In the second part of the paper, we analysed possible explanations for these trends. Looking first at disability benefits, we find Sweden saw falls in both employment rates and disability benefit recipiency, while the USA and possibly Austria saw rises in both. It is also true that Sweden reformed both the generosity and activation of its disability benefits system, whereas one of the only countries that implemented fewer reforms than Austria is the USA, whose system is effectively unchanged over this period (the only other example of this being Italy). However, the trends diverged elsewhere. Health employment gaps were stable in Denmark, Germany, Switzerland, and England, but disability benefit receipt rose in Germany and fell in the other countries. Moreover, this stability in health employment gaps belies the substantial efforts in all these countries to reform their disability benefits in the same direction as Sweden (see Figure 1). Non-negligible proportions of claimants also work in several countries, and clearly we cannot expect a reduction in their benefit claims to improve employment.

Finally, we examined trends in labour market factors relevant to the employment of older workers in poor health: part-time work, job tenure, and recruitment and retention. In none of these cases do we find convincing explanations for the USA's exceptional employment trend. Employment trends over time seem to be primarily driven by trends in full-time work rather than shifts between other working categories (notwithstanding that some countries did see significant trends in part-time work among older workers in poor health). Despite considerable variations in job tenure between countries, there are few signs of increasing/decreasing health gaps in job tenure over time. And while only some countries have seen rises in the retention rate of older workers across survey waves (for 2012/3→2014/5 compared to 2004/5→2006/7), the main differences between the USA and other countries in the follow-ups waves (2014/5 vs. 2006/7) are due to differences in the baseline employment rates (in 2012/13 vs. 2004/5) rather than in the between-wave transition probabilities.

In summary, one interpretation of our findings is that this highlights the role of disability benefit reform: the uniqueness of the USA is more apparent in its (lack of) disability benefit reform than in specific changes in the labour market. Yet we must bear in mind that the widespread adoption of disability benefit reforms has not reduced the employment gap between older workers in better vs. worse health in most countries. Assuming that these results are robust, there are several substantive interpretations, two of which fit the conventional wisdom. Firstly, what appears as similar reform trajectories in Figure 1 conceals substantial variation in policy detail and implementation, and it may be the case that many reforms were simply ineffective. Secondly, without disability benefit reform, the health employment gap in many countries may have deteriorated: perhaps rising employment rates among older workers would otherwise have been concentrated among those in good health (for example, if there are increasing capacities and incentives to work longer that are not always available for those in poor health).

A third interpretation is however possible, that the conventional wisdom about the benefits of restricted eligibility and incentives are wrong. There are several ways in which robust policy evaluations may nevertheless be misleading about the overall effect of policy reforms in the long-term, for example due to competition or spillover effects that can lead to interventions having long-term impacts even on control groups. For example, econometric evaluations suggest that restricted eligibility/incentives will lead to short-term employment gains due to increased incentives (e.g. de Jong, et al., 2011, Staubli, 2011), but van der Wel et al (2012) imply that reductions in benefit generosity will in the long-run undermine wider social investment strategies. Short-term gains may be offset in the long-term through deteriorating support and employer practice, which may ultimately lead to lower employment for those in poor health – as van der Wel et al show, we generally find the highest employment rates in the Nordic countries, which have relatively generous benefit systems alongside substantial investment in integration.

Whichever of these interpretations is proved to be correct in future research, it is already clear that the presumed improvements in the health employment gap have not materialised in many countries – and that the unique situation in the USA in particular deserves further attention.

Competing interests statement

The authors have no other competing interests to declare.

Acknowledgements

Data obtained through Gateway to Global Aging Data, produced by the USC Program on Global Aging, Health & Policy, with funding from the National Institute on Aging (R01 AG030153).

ELSA data were made available through the UK Data Archive. ELSA was developed by a team of researchers based at the NatCen Social Research, University College London and the Institute for Fiscal Studies. The data were collected by NatCen Social Research. The funding is provided by the National Institute of Aging in the United States, and a consortium of UK government departments co-ordinated by the Office for National Statistics. The developers and funders of ELSA and the Archive do not bear any responsibility for the analyses or interpretations presented here.

The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

This paper uses data from SHARE Waves 1, 2, 3 (SHARELIFE), 4, 5 and 6 (DOIs: 10.6103/SHARE.w1.611, 10.6103/SHARE.w2.611, 10.6103/SHARE.w3.611, 10.6103/SHARE.w4.611, 10.6103/SHARE.w5.611, 10.6103/SHARE.w6.611), see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Ad-

ditional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

References

1. Alavinia, S.M., Burdorf, A., 2008. Unemployment and retirement and ill-health: a cross-sectional analysis across European countries. International archives of occupational and environmental health 82**,** 39-45.
2. Angelini, V., Cavapozzi, D., Paccagnella, O., 2011. Dynamics of reporting work disability in Europe. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174**,** 621-638.
3. Avendano, M., Glymour, M.M., Banks, J., Mackenbach, J.P., 2009. Health Disadvantage in US Adults Aged 50 to 74 Years: A Comparison of the Health of Rich and Poor Americans With That of Europeans. American Journal of Public Health 99**,** 540-548.
4. Banks, J., Smith, J.P., 2012. International Comparisons in Health Economics: Evidence from Aging Studies. Annual Review of Economics 4**,** 57-81.
5. Baumberg, B., Jones, M., Wass, V., 2015. Disability prevalence and disability-related employment gaps in the UK 1998–2012: Different trends in different surveys? Social Science & Medicine 141**,** 72-81.
6. Beaumaster, S., Chien, S., Lau, S., Lin, A., Phillips, D., Wilkens, J., Lee, J., 2017. Harmonized SHARE Documentation Version D.2, Gateway to Global Aging Data.
7. Benitez-Silva, H., Disney, R., Jiménez-Martín, S., 2010. Disability, capacity for work and the business cycle: an international perspective. Economic Policy 25**,** 483-536.
8. Böheim, R., Leoni, T., 2018. Sickness and disability policies: Reform paths in OECD countries between 1990 and 2014. International Journal of Social Welfare 27**,** 168-185.

Börsch-Supan, A. (2018). Survey of Health, Ageing and Retirement in Europe (SHARE) Waves 1-6. Release version: 6.1.0.

9. Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S., 2013. Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). International Journal of Epidemiology 42**,** 992-1001.
10. Breeze, E., Lang, I., 2006. Physical functioning in a community context. In: Banks, J., and et al, (Eds.), Living in the 21st Century: older people in England (The 2006 English Longitudinal Study of Ageing, Wave 3). Institute for Fiscal Studies, London, pp. 57-117.
11. Bryan, M.L., Jenkins, S.P., 2016. Multilevel Modelling of Country Effects: A Cautionary Tale. European Sociological Review 32**,** 3-22.
12. Bugliari, D., Campbell, N., Chan, C., Hayden, O., Hurd, M., Main, R., Mallett, J., McCullough, C., Meijer, E., Moldoff, M., Pantoja, P., Rohwedder, S., St.Clair, P., 2016. RAND HRS Data Documentation, Version P, RAND Center for the Study of Aging.

13. Burkhauser, R.V., Daly, M.C., McVicar, D., Wilkins, R., 2014. Disability benefit growth and disability reform in the US: lessons from other OECD nations. IZA Journal of Labor Policy 3**,** 1-30.
14. Busemeyer, M.R., 2005. Pension reform in Germany and Austria: system change vs. quantitative retrenchment. West European Politics 28**,** 569-591.
15. Cieza, A., Oberhauser, C., Bickenbach, J., Jones, R.N., Üstün, T.B., Kostanjsek, N., Morris, J.N., Chatterji, S., 2015. The English are healthier than the Americans: really? International Journal of Epidemiology**,** 229-238.
16. Crimmins, E.M., Kim, J.K., Solé-Auró, A., 2010. Gender differences in health: results from SHARE, ELSA and HRS. European Journal of Public Health 21**,** 81-91.
17. Croda, E., Skinner, J., Yasaitis, L., Lindeboom, M., 2015. The Health of Disability Insurance Enrollees: An International Comparison.
18. de Jong, P., Lindeboom, M., van der Klaauw, B., 2011. Screening Disability Insurance Applications. Journal of the European Economic Association 9**,** 106-129.
19. Erosa, A., Fuster, L., Kambourov, G., 2012. Labor supply and government programs: A cross-country analysis. Journal of Monetary Economics 59**,** 84-107.
20. Franche, R.-L., Cullen, K., Clarke, J., Irvin, E., Sinclair, S., Frank, J., et al, 2005. Workplace-Based Return-to-Work Interventions: A Systematic Review of the Quantitative Literature. Journal of Occupational Rehabilitation 15**,** 607-631.
21. Gallie, D., 2009. Employment regimes and the quality of work. Oxford University Press, Oxford.
22. Geiger, B.B., 2017. Benefits conditionality for disabled people: Stylized facts from a review of international evidence and practice. Journal of Poverty and Social Justice 25**,** 107-128.
23. Geiger, B.B., van der Wel, K.A., Tøge, A.G., 2017. Success and failure in narrowing the disability employment gap: comparing levels and trends across Europe 2002–2014. BMC Public Health 17**,** 928.
24. Hellevik, O., 2009. Linear versus logistic regression when the dependent variable is a dichotomy. Quality & Quantity 43**,** 59-74.
25. Hogelund, J., 2003. In search of effective disability policy: comparing the developments and outcomes of the Dutch and Danish disability policies. Amsterdam Univ. Press, 2003, Amsterdam.

Health and Retirement Study, 2018. RAND HRS HRS Longitudinal File 2014 (V2) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.

26. Jones, M.K., 2006. Is there employment discrimination against the disabled? Economics Letters 92**,** 32-37.
27. Jones, M.K., 2007. Does part-time employment provide a way of accommodating a disability? Manchester School 75**,** 695-716.
28. Jürges, H., 2007. True health vs response styles: exploring cross-country differences in self-reported health. Health Economics 16**,** 163-178.
29. Kalwij, A., Vermeulen, F., 2008. Health and labour force participation of older people in Europe: what do objective health indicators add to the analysis? Health Economics 17**,** 619-638.
30. Livermore, G.A., Honeycutt, T.C., 2015. Employment and Economic Well-Being of People With and Without Disabilities Before and After the Great Recession. Journal of Disability Policy Studies 26**,** 70-79.

Marmot, M., Oldfield, Z., Clemens, S., Blake, M., Phelps, A., Nazroo, J., Steptoe, A., Rogers, N., Banks, J., Oskala, A. (2018). English Longitudinal Study of Ageing: Waves 0-8, 1998-2017. [data collection]. 28th Edition. UK Data Service. SN: 5050, http://doi.org/10.5255/UKDA-SN-5050-15

31. Mitra, S., Kruse, D., 2016. Are workers with disabilities more likely to be displaced? The International Journal of Human Resource Management 27**,** 1550-1579.
32. Morris, Z., 2016. Constructing the need for retrenchment: disability benefits in the United States and Great Britain. Policy & Politics 44**,** 609-626.

33. OECD, 2003. Transforming Disability into Ability: Policies to Promote Work and Income Security for Disabled People. OECD, Paris.
34. OECD, 2010. Sickness, Disability and Work: breaking the barriers. A synthesis of findings across OECD countries. OECD, Paris.
35. Phillips, D., Lin, Y.-C., Wight, J., Chien, S., Lee, J., 2017. Harmonized ELSA Documentation, Version E, Gateway to Global Aging Data.
36. Poterba, J., Venti, S., Wise, D.A., 2013. Health, education, and the postretirement evolution of household assets. Journal of Human Capital 7**,** 297-339.
RAND, 2018. RAND HRS Longitudinal File 2014 (V2). Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA, February 2018.
37. Robroek, S.J., Schuring, M., Croezen, S., Stattin, M., Burdorf, A., 2013. Poor health, unhealthy behaviors, and unfavorable work characteristics influence pathways of exit from paid employment among older workers in Europe: a four year follow-up study. Scandinavian journal of work, environment & health**,** 125-133.
38. Scharle, Á., Váradi, B., Samu, F., 2015. Policy convergence across welfare regimes: the case of disability policies, WWWforEurope Working Paper No. 76. WWWforEurope
39. Schuring, M., Burdorf, L., Kunst, A., Mackenbach, J., 2007. The effects of ill health on entering and maintaining paid employment: evidence in European countries. Journal of Epidemiology and Community Health 61**,** 597-604.
40. Sonnega, A., Faul, J.D., Ofstedal, M.B., Langa, K.M., Phillips, J.W.R., Weir, D.R., 2014. Cohort Profile: the Health and Retirement Study (HRS). International Journal of Epidemiology 43**,** 576-585.
41. Staubli, S., 2011. The impact of stricter criteria for disability insurance on labor force participation. Journal of Public Economics 95**,** 1223-1235.
42. Steptoe, A., Breeze, E., Banks, J., Nazroo, J., 2013. Cohort Profile: The English Longitudinal Study of Ageing. International Journal of Epidemiology 42**,** 1640-1648.
43. Trevisan, E., Zantomio, F., 2016. The impact of acute health shocks on the labour supply of older workers: Evidence from sixteen European countries. Labour Economics 43**,** 171-185.
44. Van Dalen, H.P., Henkens, K., Wang, M., 2014. Recharging or retiring older workers? Uncovering the age-based strategies of European employers. The Gerontologist 55**,** 814-824.
45. van der Wel, K.A., Dahl, E., Thielen, K., 2012. Social Inequalities in "Sickness": Does Welfare State Regime Type Make a Difference? A Multilevel Analysis of Men and Women in 26 European Countries. International Journal of Health Services 42**,** 235-255.
46. Whitehead, M., Clayton, S., Holland, P., Drever, F., Barr, B., Gosling, R., Dahl, E., Van Der Wel, K.A., Westin, S., Burström, B., Nylen, L., Lundberg, O., Diderichsen, F., Thielen, K., Ng, E., Uppal, S., Chen, W.-H., 2009. Helping chronically ill or disabled people into work: what can we learn from international comparative analyses?, Final report to the Public Health Research Programme, Department of Health.
47. Wise, D.A., 2017. Social Security Programs and Retirement Around the World: The Capacity to Work at Older Ages. University of Chicago Press for the National Bureau of Economic Research (NBER), Chicago.
48. World Bank, WHO, 2011. World Report on Disability. World Health Organization (WHO), Geneva.

APPENDICES FOR
The growing American health penalty: International trends in the employment of older workers with poor health

Table of contents for Appendices:

1.Further details on HRS/ELSA/SHARE data

Use of the Gateway to Global Aging Data RAND Harmonized dataset

Full detail of the Gateway to Global Aging Data (G2Aging)-constructed variables are given in the G2Aging codebooks.[9] We summarise key points here, alongside full details of further changes that we made to enhance comparability.

Employment variables

The employment variable in the G2Aging harmonized dataset is not comparable across surveys or over time. Most of the questions are roughly similar to one another: in HRS respondents were asked if they were doing work 'at the present time', in ELSA they were asked about working in the 'last month', and new interviewees in SHARE were asked about working in the past 4 weeks (for new interviewees). However, repeat interviewees in SHARE were instead asked if they had worked since the last interview, usually about two years previously.

It is however possible to make these questions comparable across time and surveys. To do this, we used information given in SHARE by repeat interviewees on their employment history since the last wave to create a measure of whether or not they had worked in the past month. This is then comparable to new SHARE interviewees and also to the ELSA/HRS measures. (Note that because only the month and not the full date of job starts or terminations are given, this strictly refers to employment in the present or past *calendar* month, rather than the past 30 days.)

[9] These are:

- RAND HRS Data Documentation Version P, by Delia Bugliari et al, August 2016. Labor & Population Program of the RAND Center for the Study of Aging.
- Harmonized ELSA dataset and Codebook, Version E as of April 2017 developed by the Gateway to Global Aging Data;
- Harmonized SHARE dataset and Codebook, Version D.2 as of September 2017 developed by the Gateway to Global Aging Data.

(Gateway to Global Aging Data also request acknowledgement of their funding by the National Institute on Ageing (R01 AG030153, RC2 AG036619, R03 AG043052, R03AG043052).

We also use two further employment-related measures taken from the G2Aging harmonized file: (i) working hours, and (ii) current job tenure. In both cases we restrict our analysis to those who are presently employed using the measure just described. (While Zamarro and Lee, 2011perform international comparisons using HRS-SHARE-ELSA using self-defined labour market status, we avoid this as it depends upon culturally- and temporally-variable understandings of what 'employment', 'unemployment', 'disability', etc. mean).

Disability benefit variables

While the G2Aging harmonized file does include a nominally comparable disability benefit variable, there are two errors in this variable that we correct for our analyses: (i) there are some coding errors (for income sources that are not truly disability benefits in ELSA & SHARE); (ii) the supplied derived variables do not make the SHARE variables consistent across different question versions in different waves. The supplied variables also refer to any disability benefit receipt in the past year, rather than present receipt.

We therefore use a revised version of the G2Aging variables, which are based on the same underlying questions, but which calculate the derived variables independently, as we now describe for each survey in turn.

HRS

For HRS, we use the same underlying questions in the RAND HRS Income and Wealth file. Unlike the main RAND HRS file, however, we create a dummy variable only for those who claimed disability benefits for *all* of the past year, to ensure that the measure captures people currently claiming disability benefits (a particularly important step given the analyses of simultaneous work-plus-benefits in Part 4 of the paper).[10]

[10] In HRS it would in fact be possible to include a measure of current disability benefit receipt, but this is not available in SHARE, and for consistency we have instead focussed on full-year claims.

SHARE

The disability benefit variable in the G2Aging harmonized dataset is based upon derived variables supplied by the SHARE team in their generated variables dataset, which combine underlying questions (on receipt of multiple different sources of income in the past year) into income categories. There are however two problems here:

- The main questionnaire (in English) changes between waves. For example, in wave 1 respondents are asked about 'Public disability insurance' and 'Public invalidity or incapacity pension', whereas in waves 2-5 respondents are asked about 'Main public disability insurance pension, or sickness benefits' and 'Secondary public disability insurance pension, or sickness benefits'. Moreover, at wave 6 respondents were asked separately about sickness benefits and disability benefits.

- The translated versions of these categories differ between countries. For example, sickness benefits were explicitly mentioned in waves 2-5 in Austria, Germany, Spain, Sweden, and Denmark, but not elsewhere.

The variables we use therefore aim for consistency within a country over time, rather than across countries (where sickness benefits are only inconsistently included), as follows:

Country	**Includes sick pay?**	**Inconsistencies**
Austria	Y	Sick pay not available w1, but included w2-6.
Germany	Y	Sick pay not available w1, but included w2-6.
Sweden	Y	
Spain	Y	Sick pay not available w1, but included w2-6.
Italy	N	
France	N	

Denmark	Y
Switzerland	N
Belgium	N

The full syntax used to generate the disability benefit variables is given in the replication file.

ELSA

The disability benefit variable in the G2Aging harmonized file combines all disability-related benefits together, including those that are not work-related and which in most other countries would not form part of the benefits system. This includes the extra-costs benefits Disability Living Allowance/Personal Independence Payment and Attendance Allowance, as well as Carer's Allowance. In recent waves, the G2Aging data also mistakenly misses out claims of the latest version of the out-of-work disability benefit (Employment and Support Allowance). We therefore created new variables that focused on all out-of-work disability benefits: Incapacity Benefit/Employment and Support Allowance/Severe Disablement Allowance.

Health variables

The health variables are as follows:

Self-reported global health

We included the G2Aging measure of global self-reported health, which was asked consistently across surveys (barring ELSA wave 3, which we exclude from analysis). This was dichotomised into fair/poor vs. excellent/very good/good health.

Motor skills

We used 10 binary measures of motor skills: walking 100m/one block, lifting 5kg, pulling/pushing large objects, climbing one flight of stairs, climbing several flights of stairs, stooping/kneeling/crouching, picking up a small coin, sitting for 2hrs, getting up from a chair, reaching above shoulder height. These are sometimes referred to as 'Nagi functions' (Crimmins, et al., 2010), and which refer to *'movements involving the upper and/or lower limbs, most of which require a degree of muscle strength but a few of which are more to do with dexterity and flexibility'* (Breeze and Lang, 2006).

While our analysis only requires the assumption that these variables are comparable over time within each country, for ease of interpretation we improve their cross-national comparability in several ways. Firstly we assume that anyone who struggles walking up one flight of stairs would struggle with several flights of stairs (an assumption made by question filtering in HRS, but the questions were asked independently in ELSA/SHARE, and about 1% of respondents who report no problems with several flights of stairs report problems with a single flight). Secondly, a small proportion of HRS (but not ELSA/SHARE) responses are coded as 'don't do', which was not listed on the response options but was allowed as a spontaneous response. For most motor skills, we assumed that 'don't do' means that they cannot do the task for health reasons, because these tasks are effectively universal. However, for walking 100m or climbing stairs, we have treated 'don't do' as missing, as it seems plausible that some HRS respondents genuinely do not do these tasks.

Functional disability

We used two measures of functional disability, one for any limitation in Activities of Daily Living (ADLs), one for any limitation in Instrumental Activities of Daily Living (IADLs). As we note in the main text, *'The former are basic and universal physical tasks such as eating, the latter are mixtures of physical and cognitive competencies such as preparing a hot meal (Breeze and Lang, 2006)'*. The full list of measures underlying the scales are problems with:

ADLs	**IADLs**

Dressing	Preparing a hot meal
Walking across a room	Shopping for groceries
Bathing or showering	Making telephone calls
Eating	Taking medications
Getting in or out of bed	Managing money
Using the toilet	

As for motor skills, a small proportion of HRS (but not ELSA/SHARE) responses are coded as 'don't do'. We generally assumed that 'don't do' means that they cannot do the task for health reasons; but for the IADLs we used HRS's follow-up question on the reasons for not doing the activity.

Doctor-diagnosed chronic disease

We use seven measures of self-reported lifetime doctor-diagnosed chronic diseases that are available in the G2Ageing harmonized file: high blood pressure, stroke, diabetes/high blood sugar, chronic lung disease (excluding asthma), heart problems, and arthritis.

Body-Mass Index

We use two measures of non-optimal Body-Max Index (BMI) – underweight and overweight – based on self-reported height and weight in HRS & SHARE, and measured height and weight in ELSA. Because the ELSA variable therefore has more missing responses (particularly as it was only collected in alternate waves), in our main analyses we use the temporally closest BMI observation for all respondents that have a valid BMI at any wave. This substantially increases the sample size and thereby the power of the analyses, but it does this at the cost of potential biases; we therefore conduct two sensitivity analyses, one that excludes BMI completely, and one that imputes BMI within a full multiple imputation analysis, as we detail below.

Mental health

Despite claims to the contrary (e.g. Riumallo-Herl, et al., 2014) and a common focus on recent negative feelings, the mental health measure in ELSA/HRS is not comparable to the measure in

SHARE (Courtin, et al., 2015, Zamarro, et al., 2008). However, our analysis does not depend on cross-national comparability, but only the weaker assumption of comparability within countries over time – a much more plausible assumption in this case.

The mental health measure in SHARE is the 12-item Euro-D scale, while the mental health measure in HRS/ELSA is the 20-item CES-D scale. We use two versions of these scales in our analysis: a normalized continuous version, and a binary measure of poor mental health based on the standard cut-off for the scale measure (3+ in ELSA/HRS, 4+ in SHARE); see Crimmins et al (2010) and Courtin et al (2015) for a detailed description of these measures.

2.Missing data

The following table presents the prevalence of missing data by country-wave (using unweighted data), dividing between: (i) no missing data; (ii) just BMI missing (which is particularly the case for England); (iii) one other variable missing (primarily mental health/ADLs/mobility in the USA); and (iv) 2+ variables missing:

Table A1: ***Missing data by wave.***

	No missing data				***Just BMI missing***			
Country	Wave 1	Wave 2	Wave 5	Wave 6	Wave 1	Wave 2	Wave 5	Wave 6
USA	85.9	90.5	92.5	91.8	0.3	0.2	0.4	0.5
England	90.7		84.9	69.4	8.1		11.6	26.6
Austria	99.6	98.9	97.2	99.8	0.0	0.7	0.1	0.1
Belgium	99.0	98.9	99.2	98.5	0.5	0.7	0.3	0.5
Denmark	99.5	98.4	99.0	99.3	0.1	0.4	0.4	0.4
France	96.0	97.5	99.5	98.5	0.8	0.3	0.1	0.4
Germany	99.0	98.3	98.6	99.0	0.3	1.2	0.3	0.1
Greece	98.9	99.0		99.4	0.2	0.2		0.4
Italy	98.6	99.5	99.5	98.9	0.1	0.4	0.2	0.2
Netherlands	98.6	98.3	98.6		0.6	1.0	0.4	
Spain	97.7	97.6	99.2	98.5	1.7	1.7	0.6	0.8
Sweden	99.7	97.7	98.9	98.2	0.2	0.7	0.6	0.3
Switzerland	99.6	99.3	100.0	99.8	0.4	0.2	0.0	0.0

	One other variable missing				***2+ variables missing***			
Country	Wave 1	Wave 2	Wave 5	Wave 6	Wave 1	Wave 2	Wave 5	Wave 6
USA	12.2	8.2	5.9	6.7	1.7	1.2	1.3	1.1

England	0.3		0.2	0.4	1.0		3.3	3.6
Austria	0.0	0.2	0.1	0.0	0.4	0.2	2.7	0.1
Belgium	0.2	0.3	0.3	0.5	0.4	0.2	0.2	0.4
Denmark	0.1	1.0	0.4	0.4	0.3	0.2	0.2	0.0
France	0.7	0.8	0.1	0.4	2.5	1.4	0.3	0.8
Germany	0.4	0.4	0.8	0.8	0.3	0.1	0.4	0.1
Greece	0.6	0.5		0.1	0.3	0.4		0.1
Italy	0.1	0.0	0.1	0.7	1.1	0.1	0.2	0.2
Netherlands	0.1	0.4	1.0		0.8	0.3	0.1	
Spain	0.2	0.3	0.1	0.8	0.5	0.4	0.1	0.0
Sweden	0.1	1.4	0.3	1.1	0.0	0.2	0.2	0.4
Switzerland	0.0	0.1	0.0	0.2	0.0	0.4	0.0	0.1

3.Descriptive statistics

Age and gender by country-wave

Descriptive statistics on sample size, retirement age, respondents' age, and gender are given below. Note that the sample is defined as being between age 50 and the country's retirement age (see main text), and hence the gender balance of the effective sample will depend on the respective retirement ages of men and women in the sample.

Table A2: ***Distribution of age, gender and retirement age by country-wave.***

	Sample size		***Regular retirement age[1]***		***Mean age of respondents***				***Gender (%)***	
Country	2004-7	2012-15	2004-5	2014-15[2]	2004-5	2006-7	2012-13	2014-15	2004-7	2012-15
USA	13,640	16,181	65.2 - 66.0	66.0 - 66.8	56.9	58.1	58.8	59.9	46.4	46.9
England	7,018	5,153	60.0 - 65.0	62.0 - 65.0	57.4	57.2	56.9	58.2	58.0	56.1
Austria	1,136	2,491	60.0 - 65.0	60.0 - 65.0	56.6	57.2	56.5	57.4	58.4	56.7
Belgium	3,648	5,934	63.0 - 65.0	65.0	56.8	56.9	56.9	57.1	51.3	49.6
Denmark	2,437	4,177	65.0	65.0	56.9	57.5	57.8	57.6	49.4	49.9
France	2,401	2,768	60.0	61.3	54.8	55.0	55.9	55.0	49.0	49.1
Germany	3,101	5,409	65.0	65.3	57.4	57.4	57.3	57.8	49.2	49.2
Greece	3,051	2,253	60.0 - 65.0	62.0 - 65.0	56.6	56.6		57.3	52.3	48.7
Italy	2,413	4,257	60.0 - 65.0	63.8 - 66.3	56.6	56.6	56.8	56.6	57.1	53.2
Netherlands	3,369	2,146	65.0	65.0	56.8	57.1	56.2		49.8	

Spain	2,163	5,116	65.0	65.3	56.9	57.2	56.7	57.5	49.9	50.0
Sweden	3,139	3,052	65.0	65.3	57.1	57.4	57.8	58.1	50.4	50.8
Switzerland	1,315	2,577	63.0 - 65.0	64.0 - 65.0	56.5	57.0	57.3	57.9	53.6	50.9

***Note**: [1] Details of regular retirement age variable is given in the main paper; [2] Retirement age for Netherlands in the later period refers to 2012-13.*

4.Additional material for Section 3

Health-related employment gaps by wave, top tertile vs. bottom tertile

Table A3: Health-related employment gap by country

Country	Period	Gap (%)	95% CI
USA	2004-7	33.8	(31.3, 36.4)
USA	2012-5	41.9	(39.3, 44.5)
England	2004-7	32.0	(28.2, 35.8)
England	2012-5	28.8	(24.6, 33.1)
Austria	2004-7	21.8	(14.4, 29.2)
Austria	2012-5	32.1	(25.9, 38.3)
Belgium	2004-7	23.0	(18.9, 27.2)
Belgium	2012-5	25.1	(20.5, 29.7)
Denmark	2004-7	28.1	(23.4, 32.7)
Denmark	2012-5	28.2	(24.4, 31.9)
France	2004-7	21.3	(15.9, 26.8)
France	2012-5	25.3	(20.1, 30.5)
Germany	2004-7	23.9	(19.5, 28.3)
Germany	2012-5	26.5	(22.6, 30.4)
Greece	2004-7	11.3	(6.6, 16.0)
Greece	2012-5	12.6	(7.3, 18.0)
Italy	2004-7	16.8	(11.4, 22.2)
Italy	2012-5	19.7	(14.4, 24.9)
Netherlands	2004-7	28.4	(24.2, 32.6)
Netherlands	2012-5	25.1	(16.6, 33.7)
Spain	2004-7	32.5	(26.6, 38.4)
Spain	2012-5	31.2	(22.8, 39.6)
Sweden	2004-7	31.4	(27.1, 35.8)
Sweden	2012-5	24.0	(19.2, 28.8)
Switzerland	2004-7	15.8	(9.7, 21.9)
Switzerland	2012-5	10.9	(6.2, 15.7)

Figures are average marginal effects controlling for age and gender.

Trends in employment gaps and rates

Table A4: Changes in absolute employment level among those with poor health ('level') and the health-related employment gap ('gap') by country.

Country	Level (%)	95% CI	Gap (%)	95% CI
USA	-4.5	(-7.3, -1.8)	8.1	(4.8, 11.4)
England	5.6	(1.2, 10.0)	-3.2	(-8.6, 2.3)
Austria	8.2	(1.0, 15.3)	10.3	(0.8, 19.8)
Belgium	10.4	(6.0, 14.9)	2.1	(-3.8, 7.9)
Denmark	8.6	(3.8, 13.3)	0.1	(-5.6, 5.8)
France	4.1	(-1.9, 10.1)	4.0	(-3.3, 11.2)
Germany	8.8	(4.4, 13.1)	2.6	(-3.0, 8.3)
Greece	-1.1	(-5.9, 3.8)	1.3	(-5.3, 8.0)
Italy	10.3	(4.7, 15.8)	2.9	(-4.3, 10.1)
Netherlands	13.2	(5.5, 21.0)	-3.2	(-12.2, 5.7)
Spain	5.6	(-1.3, 12.5)	-1.3	(-11.0, 8.4)
Sweden	13.0	(7.5, 18.6)	-7.4	(-13.6, -1.3)
Switzerland	11.4	(5.4, 17.3)	-4.9	(-12.2, 2.5)

Figures are average marginal effects controlling for age and gender; trends refer to 2004-7 vs 2012-15.

Sensitivity analyses

Results of sensitivity analyses

Table A5: Sensitivity analyses for trend in the health-related employment gap.

	Main (%)		*Aged 54+(%)*		*Imputation (%)*		*Exc. BMI (%)*		*Logit (%)*	
USA	-4.5	*(-7.3 to -1.8)*	-4.0	*(-6.9 to -1.0)*	-4.5	*(-7.1 to -1.9)*	-4.3	*(-7.0 to -1.5)*	-4.3	*(-7.2 to -1.5)*
England	5.6	*(1.2 to 10.0)*	4.8	*(0.8 to 8.7)*	7.7	*(4.0 to 11.4)*	7.6	*(3.5 to 11.7)*	5.4	*(1.2 to 9.7)*
Austria	8.2	*(1.0 to 15.3)*	5.3	*(-1.0 to 11.6)*	7.8	*(0.7 to 14.9)*	8.9	*(1.8 to 16.1)*	7.5	*(0.8 to 14.1)*
Belgium	10.4	*(6.0 to 14.9)*	9.7	*(5.0 to 14.5)*	10.0	*(5.6 to 14.5)*	9.8	*(5.4 to 14.3)*	9.8	*(5.2 to 14.5)*
Denmark	8.6	*(3.8 to 13.3)*	9.9	*(4.5 to 15.3)*	8.0	*(3.3 to 12.7)*	8.4	*(3.7 to 13.1)*	8.6	*(3.8 to 13.5)*
France	4.1	*(-1.9 to 10.1)*	9.7	*(5.0 to 14.4)*	5.3	*(-0.6 to 11.3)*	6.2	*(0.2 to 12.2)*	3.7	*(-1.4 to 8.8)*
Germany	8.8	*(4.4 to 13.1)*	11.4	*(6.4 to 16.3)*	8.5	*(4.1 to 12.9)*	8.3	*(4.0 to 12.7)*	9.2	*(4.6 to 13.9)*
Greece	-1.1	*(-5.9 to 3.8)*	-1.6	*(-6.7 to 3.5)*	-0.5	*(-5.4 to 4.3)*	-0.1	*(-5.0 to 4.7)*	-1.7	*(-6.8 to 3.4)*
Italy	10.3	*(4.7 to 15.8)*	13.5	*(8.5 to 18.6)*	10.7	*(5.2 to 16.2)*	10.2	*(4.6 to 15.8)*	8.8	*(3.4 to 14.2)*
Netherlands	13.2	*(5.5 to 21.0)*	18.7	*(13.5 to 24.0)*	13.8	*(6.2 to 21.5)*	14.9	*(7.2 to 22.6)*	12.7	*(4.6 to 20.8)*
Spain	5.6	*(-1.3 to 12.5)*	5.7	*(-1.4 to 12.8)*	5.8	*(-1.0 to 12.5)*	4.8	*(-2.0 to 11.5)*	6.1	*(-1.3 to 13.6)*
Sweden	13.0	*(7.5 to 18.6)*	14.6	*(8.8 to 20.5)*	13.3	*(7.7 to 18.8)*	13.7	*(8.2 to 19.2)*	13.6	*(8.2 to 19.1)*
Switzerland	11.4	*(5.4 to 17.3)*	14.1	*(7.5 to 20.8)*	10.5	*(4.5 to 16.5)*	11.2	*(5.2 to 17.2)*	11.3	*(5.6 to 17.0)*

Table A6: Sensitivity analyses for employment trend among those in poor health.

	Main (%)		*Aged 54+ (%)*		*Imputation (%)*		*Exc. BMI (%)*		*Logit (%)*	
USA	8.1	*(4.8 to 11.4)*	7.1	*(3.5 to 10.8)*	8.4	*(5.3 to 11.6)*	7.9	*(4.6 to 11.2)*	7.2	*(3.8 to 10.6)*
England	-3.2	*(-8.6 to 2.3)*	-1.8	*(-6.8 to 3.2)*	-4.4	*(-8.9 to 0.04)*	-5.2	*(-10.3 to -0.1)*	-2.9	*(-8.4 to 2.5)*
Austria	10.3	*(0.8 to 19.8)*	14.5	*(5.6 to 23.4)*	12.3	*(2.9 to 21.8)*	8.5	*(-1.0 to 18.0)*	9.9	*(1.0 to 18.7)*
Belgium	2.1	*(-3.8 to 7.9)*	4.6	*(-2.3 to 11.4)*	2.2	*(-3.7 to 8.1)*	3.2	*(-2.6 to 8.9)*	3.3	*(-2.8 to 9.5)*
Denmark	0.1	*(-5.6 to 5.8)*	2.2	*(-4.6 to 9.0)*	1.2	*(-4.5 to 6.9)*	-0.2	*(-6.0 to 5.6)*	0.5	*(-5.3 to 6.3)*
France	4.0	*(-3.3 to 11.2)*	2.0	*(-4.3 to 8.3)*	3.4	*(-3.8 to 10.6)*	1.7	*(-5.6 to 9.1)*	5.1	*(-1.7 to 11.8)*
Germany	2.6	*(-3.0 to 8.3)*	1.9	*(-4.6 to 8.3)*	2.9	*(-2.8 to 8.5)*	4.2	*(-1.5 to 10.0)*	2.4	*(-3.5 to 8.3)*
Greece	1.3	*(-5.3 to 8.0)*	3.5	*(-3.8 to 10.9)*	0.8	*(-5.9 to 7.6)*	0.2	*(-6.6 to 6.9)*	2.1	*(-4.8 to 9.0)*
Italy	2.9	*(-4.3 to 10.1)*	4.4	*(-2.6 to 11.5)*	3.2	*(-4.0 to 10.3)*	3.8	*(-3.5 to 11.1)*	4.8	*(-2.3 to 12.0)*
Netherlands	-3.2	*(-12.2 to 5.7)*	-5.6	*(-12.7 to 1.5)*	-3.2	*(-12.1 to 5.7)*	-5.9	*(-15.1 to 3.2)*	-0.7	*(-9.9 to 8.6)*
Spain	-1.3	*(-11.0 to 8.4)*	3.1	*(-7.4 to 13.6)*	-0.4	*(-10.0 to 9.1)*	-0.2	*(-9.7 to 9.3)*	-1.8	*(-12.1 to 8.6)*
Sweden	-7.4	*(-13.6 to -1.3)*	-6.8	*(-13.4 to -0.2)*	-7.5	*(-13.7 to -1.3)*	-8.2	*(-14.3 to -2.0)*	-8.0	*(-14.1 to -2.0)*
Switzerland	-4.9	*(-12.2 to 2.5)*	-5.8	*(-14.2 to 2.7)*	-3.5	*(-11.0 to 3.9)*	-6.6	*(-14.0 to 0.7)*	-3.9	*(-11.4 to 3.5)*

Further details of multiple imputation

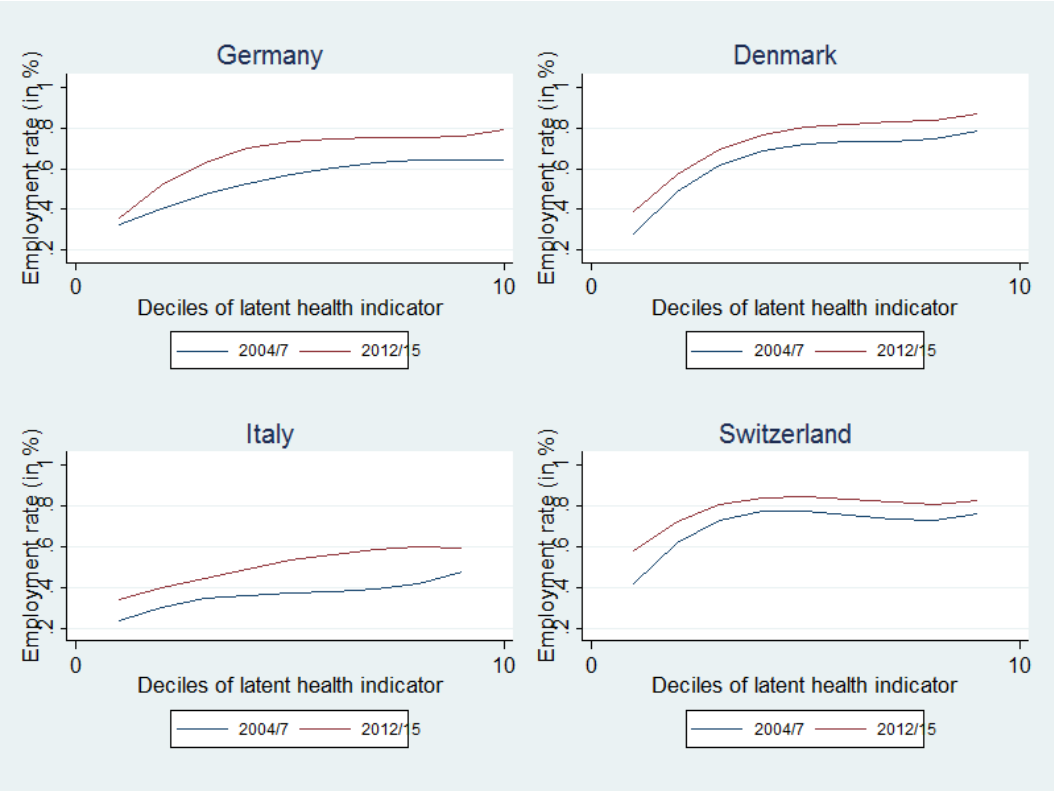
To deal with the missing data described in Online Appendix A2, we conducted multiple imputation by chained equations using the Stata multiple imputation suite. Following best practice, we constructed the imputation models to be as close to our main models as possible:

- Because we primarily treat the data as repeated cross-sections (rather than conducting longitudinal analyses at the individual level), we did not conduct a longitudinal imputation, but instead imputed responses based upon each respondent's other responses at that wave only.
- To account for country-wave clustering, imputation was done separately for each country-wave (while not necessarily being efficiency for small clusters, this is a reasonable choice if the sample sizes per cluster are large enough to make this stable; Grund, et al., 2018). The one exception was the ELSA data, for which pairs of waves were combined to enable BMI to be imputed (as BMI is only available in alternate waves).
- Imputation included all of the health variables listed in the main text, employment status, age, and gender. The only variables excluded when imputing any other variable were those that were collinear: climbing one flight of stairs vs. multiple flights of stairs; and the two ways of operationalising mental health status, both of which were excluded from each other's imputation model (while leaving both versions of each as predictors for other variables).
- Based on trace plots to check for convergence, we used 10 burn-in observations and 10 imputations.

Having imputed the data for each country-wave separately, the imputation files were combined for analysis. We then constructed our latent health measure using PCA for each imputation separately (so that 'poor health' refers to the bottom tertile within each country-wave for each imputation). We then ran analyses across the 10 imputed datasets that take into account the imputation-related uncertainty using Rubin's rules, using the Stata multiple imputation suite. Full syntax for the multiple imputation is given in our replication files.

Distribution of employment along the whole health distribution

Figure A1: Employment across the whole distribution of health, 2004-7 to 2012-15 (remaining countries).



Figures show kernel-weighted local polynomial smoothing of a regression of employment on health, controlling for age and gender.

5.Additional material for Section 4

Trends in disability benefit receipt

The table corresponding to Figure 5 in the text is as follows:

Table A7: ***Trends in disability benefit receipt across the whole older working-age population***

Country	***Trend***	***95% CI***
USA	1.5%	(0.6%, 2.4%)
England	-3.4%	(-4.7%, -2.1%)
Austria	2.4%	(0.1%, 4.7%)
Belgium	1.5%	(0.0%, 3.1%)
Denmark	-2.9%	(-4.9%, -1.0%)
France	1.5%	(0.3%, 2.7%)
Germany	2.1%	(0.7%, 3.5%)
Italy	-2.7%	(-4.2%, -1.2%)
Netherlands	-1.2%	(-4.0%, 1.7%)
Spain	-2.1%	(-4.4%, 0.1%)
Sweden	-8.5%	(-11.2%, -5.8%)
Switzerland	-2.3%	(-4.4%, -0.3%)

Trends in disability benefit receipt in Sweden

In the main text, we see a particularly sharp decline in disability benefit claims in Sweden. We carefully examined the data to check that this is not an artefactual trend due to changing questions in SHARE, and this does not appear to be the case:

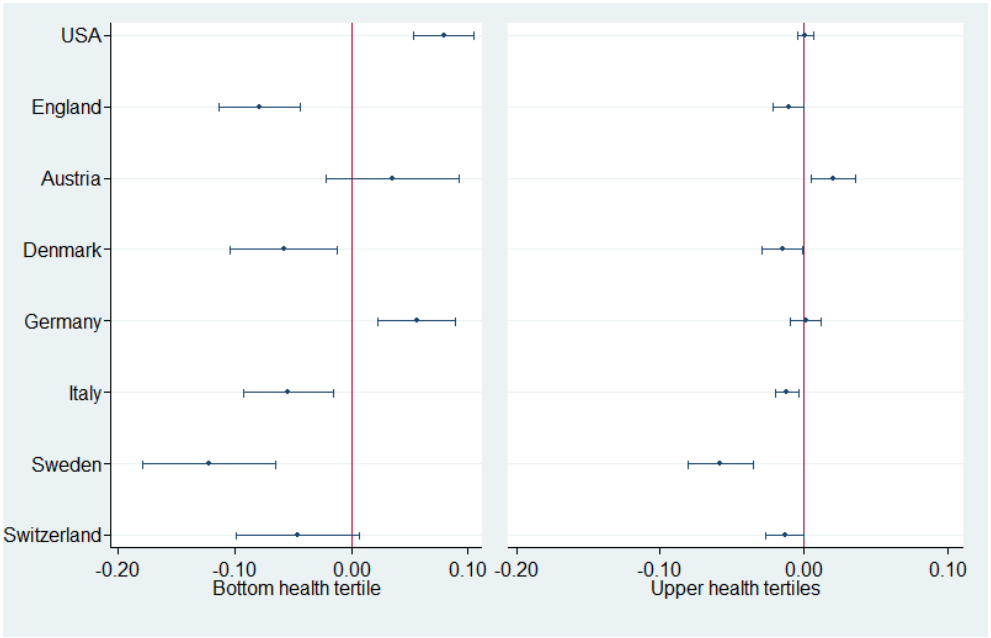
- Reported claim rates in Sweden in 2004 and 2006-7 are very similar, so the trend is not due to any changes in SHARE question wording between wave 1 and wave 2.

- OECD claim data (taken from the OECD SOC R database 2007- and from OECD (2010) for 2004-7) show an almost identical proportion decrease in disability benefit claims among the whole Swedish population using administrative data (for 2013 vs. 2004, a 28% relative decline in SHARE for people between 50 and the retirement age, and a 31% decline in the administrative data for all ages). It is however worth noting that the timing of this decline is slightly different in the SHARE data than the administrative data (in SHARE the change occurs between 2011 and 2013, while in the administrative data it occurs between 2007 and 2012).

It is also worth noting here that because of estimation problems for these logit models, we control for age, age^2, and gender using constant coefficients across all countries (rather than country-specific coefficients as in the models in the rest of the text). We have however re-run these models using OLS rather than logit models (for which there are no estimation problems in using country-specific coefficients), and these produce virtually identical results.

Trends in disability benefit receipt split by health tertile

Figure A2: Disability benefit claim rates split by health, 2004-5 to 2014-15.

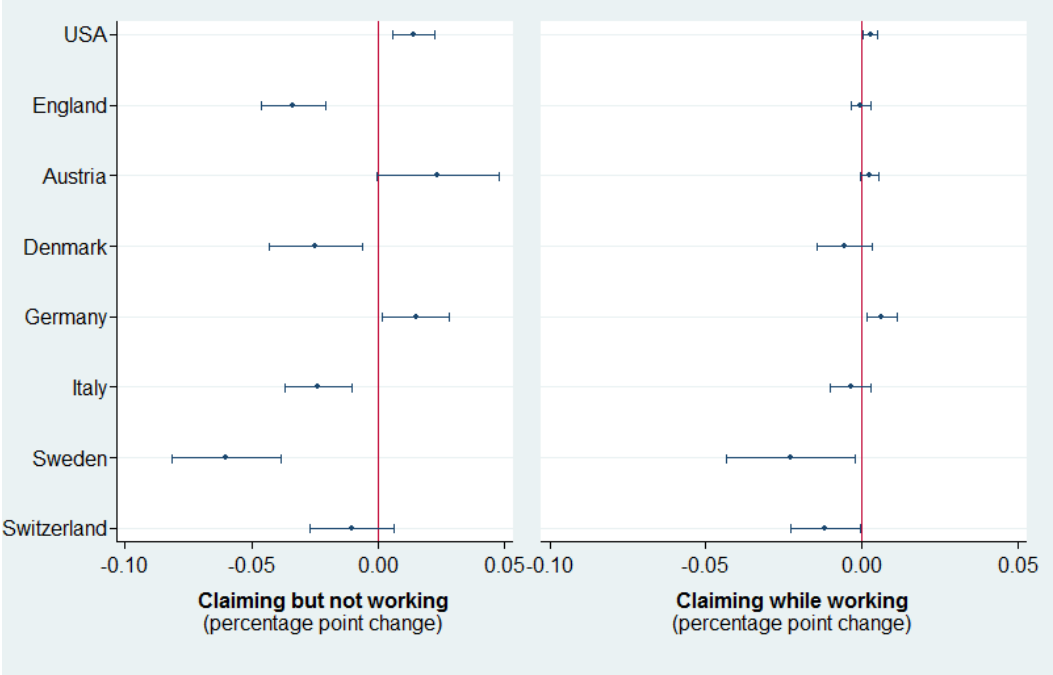


Figures show trends among the bottom health tertile (left panel) and the other two health tertiles (right panel) of our health index.

Trends in disability benefit receipt split by employment status

Figure A3 below sheds light on how DI benefit enrolment changed with time, distinguishing between working and nonworking claimants. In all countries that reduced disability benefit claims, the reduction was stronger among inactive recipients than among those who worked. This is particularly true of Sweden, where benefit enrolment rates decreased by approximately 7 percentage points for nonworking persons. Switzerland is the only exception in this respect, as the decrease in benefit rates was approximately equal for both groups. In contrast, the increase in DI benefit rates observed in Austria, Germany, and the United States was basically concentrated on higher shares of non-working benefit claimants.

Figure A3: Changes in benefit claim rates split by employment status, 2004-5 to 2014-15.



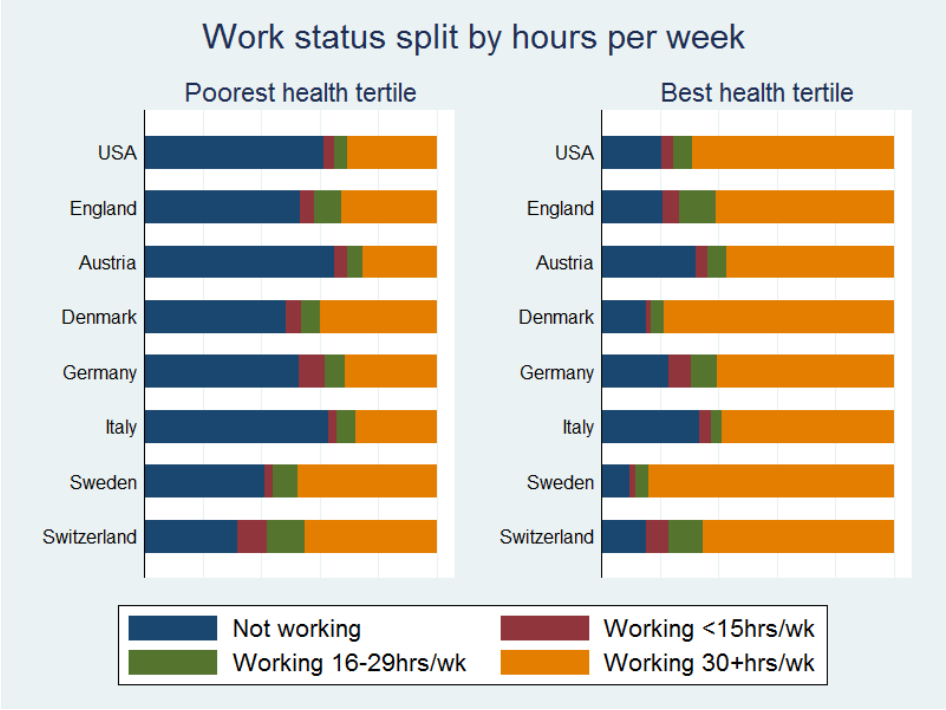
6.Additional material for Section 5

Hours of work

Comparison of hours of work by health status, pooled across waves

In all countries, the share of workers who work fewer than 30 hours per week is greater among those with poor health status than among the healthier workers. This is however true only if we base our comparison on the number of workers who are employed. As Figure A4 below shows, among the total sample of working-aged persons there is (surprisingly) little variation in part-time employment between health tertiles within countries. It is interesting to note that Sweden – which has the largest shares of people combining benefit claim and paid work activity – does not have high numbers of workers with reduced working hours.

Figure A4: Work status split by hours per week worked, all waves.



Trends in hours of work

The table below corresponds to Figure 6 in the main text:

Table A8: Trend in hours of work by country, 2004-7 to 2012-15

	Not working		***Working <15hrs/wk***		***Working 15-29hrs/wk***		***Working 30+ hrs/wk***	
Country	*Trend (%)*	*95% CI*	*Trend (%)*	*95% CI*	*Trend (%)*	*95% CI*	*Trend (%)*	*95% CI*
USA	5.5	(2.6, 8.3)	0.6	(-0.4, 1.7)	-0.6	(-1.8, 0.5)	-5.6	(-8.4, -2.9)
England	-5.6	(-9.9, -1.4)	-0.8	(-2.9, 1.4)	0.1	(-2.9, 3.1)	5.5	(1.8, 9.3)
Austria	-6.8	(-13.0, -0.5)	0.4	(-2.6, 3.3)	3.9	(1.1, 6.7)	2.9	(-2.8, 8.5)
Denmark	-8.3	(-13.1, -3.6)	0.7	(-1.5, 2.9)	0.7	(-1.7, 3.2)	6.7	(2.1, 11.4)
Germany	-9.5	(-13.9, -5.2)	4.4	(1.9, 6.8)	2.8	(0.6, 5.1)	2.5	(-1.6, 6.7)
Italy	-8.7	(-13.8, -3.6)	-2.5	(-4.1, -0.8)	2.2	(-0.5, 4.8)	9.0	(4.1, 13.9)
Sweden	-13.4	(-19.0, -7.7)	-0.8	(-2.3, 0.7)	-5.0	(-8.2, -1.8)	19.1	(13.5, 24.7)
Switzerland	-12.5	(-18.5, -6.5)	-2.8	(-6.8, 1.2)	4.5	(0.4, 8.7)	9.9	(3.8, 15.9)

Figures are average marginal effects controlling for age and gender.

Job tenure

Table A9: Job tenure by health (in years)

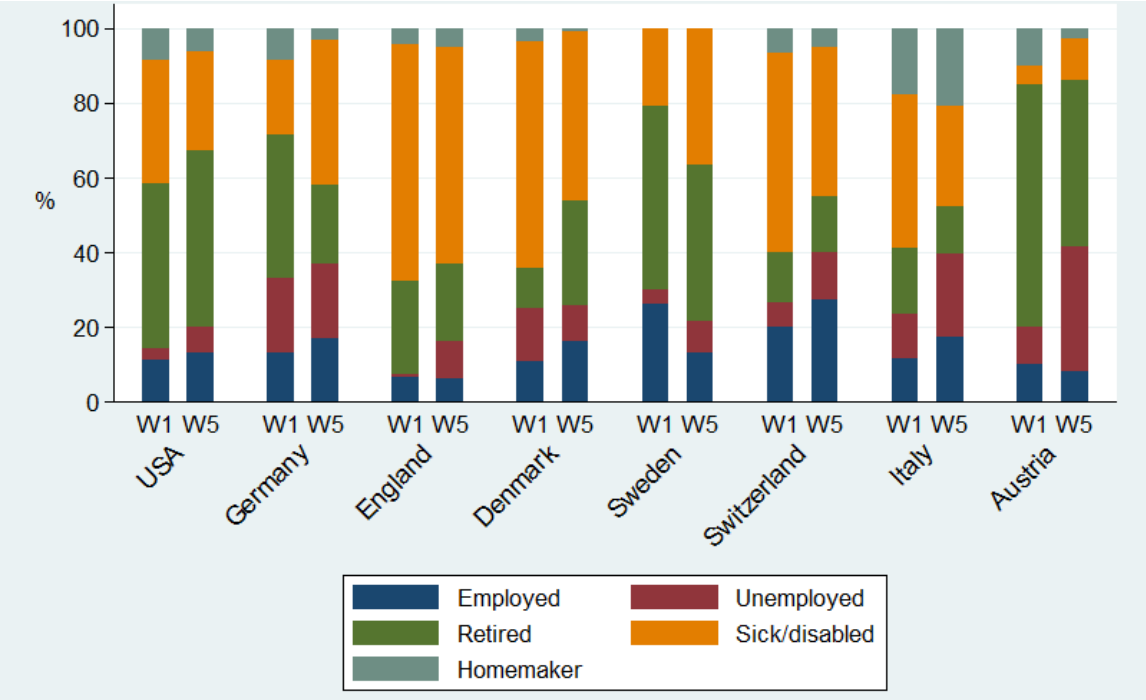
Country	*Poorest health*	*95% CI*	*Best health*	*95% CI*
USA	11.5	(11.0, 12.1)	13.0	(12.6, 13.4)
England	10.0	(9.4, 10.6)	11.4	(10.8, 12.0)
Austria	22.3	(20.7, 23.8)	21.9	(20.7, 23.1)
Belgium	24.4	(23.2, 25.7)	24.6	(23.7, 25.4)
Denmark	14.8	(14.0, 15.7)	16.1	(15.4, 16.9)
France	22.1	(21.0, 23.2)	23.0	(22.1, 23.9)
Germany	17.0	(16.1, 18.0)	19.8	(19.0, 20.6)
Greece	24.0	(22.9, 25.1)	24.7	(23.9, 25.5)
Italy	24.3	(23.0, 25.7)	27.2	(26.5, 28.0)
Netherlands	17.6	(16.1, 19.0)	18.4	(17.4, 19.4)
Spain	21.5	(19.2, 23.7)	23.5	(22.1, 24.9)
Sweden	17.2	(16.1, 18.3)	16.6	(15.6, 17.6)
Switzerland	16.8	(15.9, 17.8)	18.0	(17.1, 18.9)

Table A10: Job tenure across countries, 2012-15

Country	*Trend*	*95% CI*
USA	11.8	10.9 to 12.7
England	8.2	6.9 to 9.4
Austria	21.7	19.7 to 23.8
Denmark	14.1	13.1 to 15.1
Germany	17.1	16.1 to 18.2
Italy	23.8	22.2 to 25.3
Sweden	17.0	15.6 to 18.5
Switzerland	17.1	15.8 to 18.4

Detailed labour market transitions between waves

Figure A5: Labour market status at follow-up wave of non-workers at baseline wave (for wave 1è2 and wave 5è6)



Notes: All figures are population weighted. Homemakers and retirees in baseline (W1/W5) excluded from the sample.

References for Online Appendices

1. Breeze, E., Lang, I., 2006. Physical functioning in a community context. In: Banks, J., and et al, (Eds.), Living in the 21st Century: older people in England (The 2006 English Longitudinal Study of Ageing, Wave 3). Institute for Fiscal Studies, London, pp. 57-117.
2. Courtin, E., Knapp, M., Grundy, E., Avendano-Pabon, M., 2015. Are different measures of depressive symptoms in old age comparable? An analysis of the CES-D and Euro-D scales in 13 countries. International Journal of Methods in Psychiatric Research 24**,** 287-304.
3. Crimmins, E.M., Kim, J.K., Solé-Auró, A., 2010. Gender differences in health: results from SHARE, ELSA and HRS. European Journal of Public Health 21**,** 81-91.
4. Fernandes, M., Zamarro, G., Meijer, E., 2008. 2.2 Health Comparisons. In: Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A., Mackenbach, J., Siegrist, J., and Weber, G., (Eds.), First Results from the Survey of Health, Ageing and Retirement in Europe (2004-2007): Starting the Longitudinal Dimension. SHARE, Mannheim.
5. Grund, S., Lüdtke, O., Robitzsch, A., 2018. Multiple Imputation of Missing Data for Multilevel Models:Simulations and Recommendations. Organizational Research Methods 21**,** 111-149.
6. OECD, 2010. Sickness, Disability and Work: breaking the barriers. A synthesis of findings across OECD countries. OECD, Paris.
7. Riumallo-Herl, C., Basu, S., Stuckler, D., Courtin, E., Avendano, M., 2014. Job loss, wealth and depression during the Great Recession in the USA and Europe. International Journal of Epidemiology 43**,** 1508-1517.
8. Zamarro, G., Lee, J., 2011. Harmonization of Cross-National Studies of Aging to the Health and Retirement Study, RAND Working Paper. RAND.
9. Zamarro, G., Meijer, E., Fernandes, M., 2008. 2.3 Mental Health and Cognitive Ability. In: Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A., Mackenbach, J., Siegrist, J., and Weber, G., (Eds.), First Results from the Survey of Health, Ageing and Retirement in Europe (2004-2007): Starting the Longitudinal Dimension. SHARE, Mannheim.