Organisation for Economic Co-operation and Development

DOCUMENT CODE

For Official Use

English - Or. English

Measuring disability employment gaps:

How to get robust comparisons across countries and over time

APPENDICES

These are the Appendices of a Working Paper that addresses a key issue for research on disability issues and policies: the volatility of the measurement of disability and its impact on indicators based on this measure, such as employment rates and employment gaps.

Ben Geiger (ben.geiger@kcl.ac.uk)
Christopher Prinz (christopher.prinz@oecd.org)

Table of contents for appendices

Appendix A: How to use these methods

Appendix A1: Introduction to using these methods yourself

This Appendix guides you through the process of conducting some of the analyses in the main report for yourself on a different dataset. The full replication code is explained in Appendix A5, but most users will instead find it easier to use one of the earlier appendices, which explain how to do a specific analysis.

Please note that these Appendices are a live document, and if there are any corrections, clarifications or updates, we will try to amend these appendices on the OECD website. If you would like to be alerted to any changes, please contact Ben Geiger at ben.geiger@kcl.ac.uk.

The analyses in this report were conducted using Stata version 17. Readers using other statistical packages should be able to either follow the code, or to use AI tools to translate this into another package. If you have developed sample code for use with another package, please do contact Ben at the address above and we can make this publicly available (with appropriate accreditation) on the OECD website within these appendices.

Appendix A2: Initial estimates adjusted for age/gender

In the main text, we mention that we look at country differences in our key outcomes (disability prevalence / absolute disability employment / the disability employment gap) after controlling for age and sex, to adjust for sociodemographic differences between countries.

Rather than using raw proportions, we therefore use two regression models: a disability model (with country, age and sex as independent variables), and an employment model (with country*disability, age and sex as independent variables). We estimate marginal effects of country or country*disability in both models (holding age and sex constant at the smple mean values; see also Appendix A4), which gives us the results shown in the main report. This is relatively quick and easy to do; full syntax is given in Supplementary Code A2.

Appendix A3: Using the prevalence-adjusted disability employmet gap

The prevalence-adjusted disability employment gap, alternatively termed the Proportion Prevented from Working due to Disability (PPWD), is a simple measure – it is simply the disability employment gap multiplied by the disability prevalence rate. Like with disability prevalence and employment just above (Appendix A2), we control for age and sex to adjust for sociodemographic differences between countries – so the PPWD also comes with estimates of marginal effets following logistic regression models, except that we now multiply the estimated disability employment gap by the estimated disability prevalence

However, there is no simple way of constructing a confidence interval around this estimate. Instead, we have to 'bootstrap' the confidence intervals – that is, we treat the existing sample as if it were a population, and randomly sample from it to create 100 hypothetical samples, which are called 'replications'. In each of these, some sample members will appear once, some will appear multiple times, and some will not appear at all.

In the simplest case, the 95% confidence interval comes from the middle 95 of these 100 replications (known as 'percentile-based bootstapped confidence intervals'). Percentile-based confidence intervals have some known biases (particularly in small samples), as do bootstrapped estimates that do not take account of survey weights (Kolenikov, 2010), and it is sometimes possible to get more accurate bootstrapped confidence intervals using statistical refinements (see Hesterberg, 2014 for a clear introduction) (and Efron and Tibshirani, 1993; Hastie *et al.*, 2009 for full detail). However, in our study, we found that some of these refinements worked poorly, while the percentile-based confidence intervals work very well, and we therefore present percentile-based confidence intervals in the main study (for interested readers, see the following box).

Statistical refinements to the bootstrap

There are three sets of statistical refinements to the simplest bootstrapping approach.

Firstly, the bootstrap replications create a distribution of bootstrap estimates (in the simplest case, these are used for the percentile-based estimate). However, this distribution may be centred on a mean that differs from the full-sample estimate of the PPWD, which in bootstrapping is known as 'bias' in the bootstrap distribution (as the full-sample estimate is usually the correct one). Because the full-sample estimate is the most accurate, it is possible to use 'bias-corrected' estimates that account for this difference. These make little difference in our case, but interested users will find that these are already created in the syntax in Supplementary Code A2.

Secondly, it is possible to account for the skew of the distribution of bootstrap estimates to get more accurate estimates in the tails (which are useful for estimating the 95% confidence interval; this is known as estimates that are correct to the 'second order'). This uses 'bias-corrected and accelerated' (BCa) bootstrap estimates, which are also available in Stata. However, because of the large datasets being used here, these techniques were unusably slow compared to the main bootstrap estimates (they require 60,000 jacknife estimates within each replication, which would mean estimating each model about 25 million times) – these would have taken days or weeks on the computers we used in this analysis.

Third, there are cautions about using the bootstrap for data that requires survey weights (e.g. Hesterberg 2014). More advanced bootstrap methods do exist that account for survey weights (instead of randomly choosing people for each replication, the data are randomly reweighted in each replication; see Kolenikov 2010 and the Stata manual for *'svy bootstrap'*). We cannot benchmark this when looking at the PPWD, so instead we benchmarked it just looking at disability prevalence and disability employment, which we can calculate without using bootstrap methods. When we did this, we found the weighted method performed poorly– the confidence intervals were about one-third the size that they should have been. For this reason, we do not include the weighted bootstrap results in the main report or Supplementary Code A2, but interested readers will find this syntax in the replication materials in Supplementary Code A5.

In contrast, when we benchmark the bootstrap estimates for the disability employment gap against non-bootstrap methods, we find the unweighted bootstrap performed well (which matches wider views that percentile-based bootstrap methods are a reasonable option for large samples when BCa methods cannot be used; see Hesterberg 2014:42-49).

While it is easiest to explain bootstrapping with 100 replications, in practice we do 400 replications to reduce random noise in our estimates of the PPWD. For ease of comparison, we also use bootstrap confidence intervals for the disability employment gap in similar fashion.

The statistical code (given in Supplementray Code A2) requires us to do two steps:

1. Create a program to create the PPWD – that is, write a series of commands that create the results that we want, that are called by a single name (which is needed for bootstrapping).
2. Bootstrap the results of this program.

The outputs of this program can then be exported into Excel and turned into publication-standard tables.

Appendix A4: Using the predicted disability scale

The 'predicted disability' approach is our recommended way of creating a disability scale using impairment/activitity limitation (or other relevant) measures. This weights the measures by how strongly they predict single-item activity-limiting disability. It is similar to the health index created by Jürges (2007), though the idea of constructing an 'objective' health index in this way goes back at least to Bound (1991) – see the main text for further discussion.

Creating the predicted disability scale is straightforward – it is simply a matter of running a standard regression model, and then creating predicted probabilities based on it. To try to ensure that the impairment/activitity limitation weights reflect real associations between these measures and single-item activity-limiting disability, rather than socioeconomic patterning or differences in reporting styles, the logistic model also controls for country (similar to Jürges, 2007), employment status (to account for work-related biases in reporting disability), age (single year), gender, and education.[1] The model predicting disability is therefore:

$$y_{ij} = \beta_1 \text{limitations} + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{employment} + \beta_5 \text{education} + \beta_6 \text{country} + \varepsilon_{it}$$

…where y_{ij} is self-reported general disability for person *i* in country *j*, limitations is a vector of impairments/activity limitations measures and β_1 is a vector of regression coefficients, and β_2 to β_5 are regression coeffificients on the control terms and β_6 is a vector of regression coefficients on the country dummies.

We then use this model to estimate the predicted probability of each respondent reporting a disability, based purely on these impairment/activity limitation measures (holding age, gender, employment, education and country constant). We do this using the 'predict' function in Stata, but setting all the values of the control variables to the sample mean values. (This is helpful when using the probabilistic approach, to try to ensure that the prevalence of probabilistic disability is the same as the prevalence of single-item activity-limiting disability).[2]

[1] For some indicators, we alter the form of the health measures slightly in the predicted disability models (compared to the latent disability models). This is because there can be collinearity in the regression model that makes some weights unexpectedly negative (Lee *et al.*, 2013), which here particularly occurred for the ADL/IADL measures. We therefore combined these into measures of 'any ADL' (binary) and 'any IADL' (0/1/2+), after which all measures have positive weights ($p<0.05$) ranging from 0.25-1.10 on the logit scale (see Appendix B).

[2] Using the values of control variables at sample means produces an overall prevalence of disability in the probabilistic approach that is similar to the all-country average, unlike if we estimate marginal effects using the European Standard Population.

We also have to make a further small further adjustment to try to match the prevalence more closely, because even after this, the weighted mean of the predicted probabilities is slightly different than the weighted mean of single-item

When using a fixed cut-off on the predicted disability scale, it is usually imposible to exactly match the prevalence of disability (because the predicted disability is not smooth, instead jumping up in steps – see

Appendix A5: Using the probabilistic approach for binary disability

From a situation where we use the predicted disability scale with a fixed cut-off, there is one major difference when we move to a probabilistic scale. Rather than having a fixed threshold on the predicted disability scale, we then probabilistically assign each person to the disability category using each person's predicted probability of having a disability. In practice, this means generating a random value between 0 and 1 for each person, and then treating them as 'having a disability' if their predicted probability is greater than this random number. We do this over 400 replications where the random number if generated again, so that e.g. a person with a 10% chance of being classified as having a disability (in the predicted disability scale) will be treated as 'having a disability' in 40 of the 400 replications (on average).

We do these 400 replications using the 'simulate' command in Stata (see Supplementary Code A5). This is relatively straightforward, as shown in the supplementary code. Note that it is also possible to do these estimates using bootstrapping (the underlying programme is identical in both cases, with the main difference being that this is repeated using the 'bootstrap' than 'simulate' command – see replication file). The 'simulate' command is both more accurate and faster, but the bootstrapping approach is also valid and produces nearly-identical results (and may be preferable if you are also estimating the PPWD, which requires bootstrapping; see Appendix A3 above).[3]

---

activity-limiting disability (because of the functional form of the logistic function). We therefore rescale the predicted probability so that the weighted mean of predicted probability is the same as the mean of single-item activity-limiting disability prevalence (this can results in a small number of predicted probabilities over 1, but this does not materially affect the results). This correction is only relevant for the probabilistic disability method, and can be ignored if preferred as the difference it makes is small.

[3] If we use the 'bootstrap' command for the probabilistic approach, then we have the problem that the full sample estimate is not a good estimate of the mean value (it is but one of the many different replications for probabilistic disaibility). We therefore have three options:

1. *Just use 'simulate'*: in this approach, the probabilistic method uses the full sample, and considers the Monte Carlo error in the probabilistic disability measure across 1,000 replications. (The 95% confidence interval is obtained from the mean value of the upper and lower bounds of the 95% confidence interval across 1,000 replications).
2. *Just use bootstrapping*: this uses the bootstrapping approach described in Appendix A3, except that in each bootstrap replication, we use a different random number to probabilistically assign people to the disability category. The sampling error is therefore combined with the Monte Carlo error from whether or not someone is categorised as having a disability. Because the full-sample estimates are biased (they are just one realisation of the random error in whether people have a disability), we use the average value in the bootstrap replications as our main estimate (i.e. we unusually use the bias-corrected estimate of the mean value, alongside the percentile-based 95% CI).
3. *Use bootstrapping + simulate*: it is also possible to combine these approaches, so that we use 'simulate' to get the full-sample estimate that accounts for the randomness in the probabilistic disability measure (across 1,000 replications), and then use 'bootstrap' to get the 95% confidence interval around this that accounts for sampling error. This produces similar results to the two other approaches, but takes much longer to estimate (e.g. the bootstrapping and simulate approaches both take 135-140 minutes, but the combined bootstrapping + simulate approach takes 340 minutes).

Appendix A6: Full replication code for results in the main report

Because we are not the data owner for the SHARE-ELSA-HRS or EHIS data, we cannot share these directly. However, these are public datasets that readers should be able to obtain without charge:

- SHARE data are available from the SHARE website (https://share-eric.eu/data/data-access). ELSA data are available from the UK Data Service (Study Number 5050, https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=5050). Harmonized HRS data are available from the Health and Retirement Study website (https://hrsdata.isr.umich.edu/data-products/gateway-harmonized-hrs).
- EHIS data are available from Eurostat via the application process detailed at https://ec.europa.eu/eurostat/web/microdata/european-health-interview-survey. This requires an application process that takes roughly 2-3 months, as EU Member States are consulted on whether the proposed use of the data is reasonable (this process is however relatively straightforward, and our understanding is that the overwhelming majority of applications are granted). Given that the UK had left the EU by the time the EHIS wave 3 were published, the UK data for EHIS wave 3 is available from the UK Data Service (Study Number 7881, https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7881).

We here use the RAND harmonised versions of the datasets, which produce comparable versions of many key variables – although we go beyond this, as we explain in Appendix D. The RAND harmonised datasets are available via the Gateway to Global Aging Data web portal, https://g2aging.org/app/hrd/overview (the RAND harmonised HRS data is downloadable directly from RAND; for ELSA and SHARE, the public versions need to be downloaded, and then the RAND harmonising code needs to be run – we use version D.2 for SHARE and version E for ELSA). Before running our code that follows, you need to run this RAND harmonising code. (The data cleaning files start with the files ${elsadir}\H_ELSA.dta, ${hrsdir}\rndhrs_p.dta and ${sharedir}\H_SHARE_D2.dta, which are the files that result from running the RAND harmonising code).

Using the RAND harmonised data, the full replication code includes a series of different Stata syntax (.do) files, as follows:

Filename	
General and set-up	
0_master.do	This is the master file, which all of the other syntax files are called from.
0_globals.do	Anything that changes between analyses – e.g. filenames and locations, variable lists – is stored in a 'global' (basically, a named placeholder). This means that you just need to change these things once, in this file, and then all of the other syntax files should work as they should do (rather than having to do a find-and-replace in every single file).
0_initialising.do	This is probably not necessary for most users (it just clear the memory and ensures that the right packages are loaded) – but if you are getting an error because the code is trying to run a Stata package that you don't have, then run this once and it will load these onto your computer.
Data cleaning	
1_cleaning_ELSA.do 1_cleaning_SHARE.do 1_cleaning_HRS.do	These do further data cleaning on each file separately. This usually involves merging in additional variables from the SHARE-ELSA-HRS data that are not included in the RAND harmonised dataset.

2_cleaningXsurvey _labelling.do	This merges the ELSA, SHARE and HRS files into one working datafile, together with a small amount of further data cleaning.
Analysis	
4_disweights.do	This creates the latent and predicted disability scales, and also the various disability measures that are based on these. To ensure that this works correctly, sample restrictions (e.g. age restrictions) need to be done before running this file.
5_nonbootstrap.do	This produces the conventional disability prevalence and disability employment gap results (after controlling for age and gender; see Appendix A2).
5_results_simulate.do	This is the programme that does the simulations to enable use of probabilistic disability (see Appendix A5). Note: this code loads the programme, which is then run from within 0_master.do.
5_results_bootstrap.do	This is the programme that produces bootstrapped results for all results – this includes PPWD (see Appendix A3), the predicted disability scale with a fixed cut-off (see Appendix A4), and the alternative approach to using probabilistic disability (see Appendix A5). Note: this code loads the programme, which is then run from within 0_master.do.
5_results_ weightedbootstrap.do	This is the same as 5_results_bootstrap.do, except that it allows the use of weights (see the wider syntax within 0_master.do for how this is used). Note: as the appendices note, these results do not appear to be trustworthy (the confidence intervals are too narrow) – we have however included this in the replication file in case users want to explore this, and suggest ways of improving this syntax.

Using the EHIS data, the full replication code includes a series of different Stata syntax (.do) files, as follows:

Filename	
General and set-up	
0_master.do	This is the master file, which all of the other syntax files are called from.
0_globals.do	Anything that changes between analyses – e.g. filenames and locations, variable lists – is stored in a 'global' (basically, a named placeholder). This means that you just need to change these things once, in this file, and then all of the other syntax files should work as they should do (rather than having to do a find-and-replace in every single file).
Data cleaning	
1_prep_EHIS2.do 1_prep_EHIS3.do	These do data cleaning separately on the two waves of EHIS, EHIS wave 2 and wave 3.
Analysis	
4_disweights.do	This creates the latent and predicted disability scales, and also the various disability measures that are based on these. To

	ensure that this works correctly, sample restrictions (e.g. age restrictions) need to be done before running this file.
5_EHISandSHARE_0master.do 5_EHISandSHARE_nonbootstrap.do	These do the set-up and analysis for comparing EHIS and SHARE. Note that the data cleaning for the SHARE data is done at the bottom of the SHARE version of 0_master.do (rather than in the EHIS replication files) – you need to check that the file locations are consistent between the different syntax files.
6_allages_0master.do 6_allages_bootstrap.do	These do the set-up and analysis for doing the all-ages analysis of EHIS wave 3 (rather than just the 50-69 age group, which is used for the EHISandSHARE analysis).
7_EHIStrend_0master.do 7_EHIStrend_bootstrap.do	These do the set-up and analysis for looking at trends between EHIS wave 2 and wave 3, for all ages.

It is also worth noting the following:

- *Setting up the replication files to work on your computer:* look at the 0_globals.do file for each replication, which includes the file locations for everything that is used in these syntax files. You will need to change these file locations to wherever the data files, .do files and working folders are on your computer – but once you have done this in 0_globals.do, all of the rest of the syntax files should work smoothly.
- *Control variables:* the following control variables are used:
    - In all SHARE-ELSA-HRS analyses, the disability weights look at the association of functional impairments with single-item activity-limiting disability controlling for country, employment status, age (single year), gender, and education (see Appendix A4 above). In EHIS analyses we do the same, except we do not use education (as we judged that the education variable in EHIS is not comparable across countries).
    - In the main analyses, the regressions that compare countries in the prevalence of disability or in disability employment control for only age and gender. That is: we want to account for the different age-sex composition of the population in different countries, but not differences in employment or education.
    - In the analyses that compare two datasets (e.g. EHIS vs. ELSA-SHARE, or EHIS wave 2 vs. wave 3), we allow for the effects of country to vary across datasets, and therefore also allow for age and gender to vary across datasets (that is, we include the interaction of dataset with age and gender). However, this is only done for the disability prevalence/employment models; the models to create the predicted disability weights constrain the control variables to have constant effects across datasets. (But readers who wish to run similar analyses in future could easily allow these controls to have different effects in different datasets).

If you have any problems with running this replication file or related queries, please do not hesitate to get in touch with Ben Geiger at ben.geiger@kcl.ac.uk.

Appendix B: Further Tables/Figures

Appendix B1: Additional results

The following additional results are included in this appendix:

- Table B1: Survey differences in the disability employment rate in 50-69 year olds
- Figure B1: The difference between employment gaps of medium vs. high disability probabilities
- Table B2: Weights used in creating disability scales
- Figure B2: The combined impact of methodoligcal changes: the difference between using a probabilistic vs. single-item activity-limiting disability measure on prevalence and employment gaps among 50-69 year olds
- Figure B3: How final disability employment results differ for single-item activity-limiting disability vs. probabilistic disability, for 20-69 year olds
- Figure B4: How EHIS international comparisons change when using single-item activity-limiting disability vs. probabilistic disability based on predicted disability scales, for 20-69 year olds in 2013-15
- Figure B5: How trends in EHIS international comparisons change when using single-item activity-limiting disability vs. probabilistic disability based on predicted disability scales, for 20-69 year olds in 2013-15 vs. 2018-20

Table B1: Survey differences in the disability employment rate in 50-69-year-olds

	SHARE-ELSA		EHIS		**Difference**	
Sweden	62.4	[57.7,67.2]	63.6	[55.7,71.4]	**-1.1**	[-10.3,8.0]
Estonia	50.8	[46.6,54.9]	47.1	[41.7,52.5]	**3.7**	[-3.2,10.5]
Germany	47.1	[43.6,50.5]	43.9	[40.2,47.5]	**3.2**	[-1.8,8.2]
Denmark	45.4	[40.9,49.9]	50.6	[45.6,55.6]	**-5.2**	[-11.9,1.6]
France	32.4	[28.1,36.8]	25.7	[22.7,28.7]	**6.7**	[1.4,12.0]
Czechia	29.5	[23.1,36.0]	33.4	[29.4,37.4]	**-3.9**	[-11.4,3.7]
Italy	28.2	[23.7,32.6]	31.6	[28.8,34.4]	**-3.4**	[-8.7,1.9]
Austria	24.5	[20.9,28.2]	27.3	[24.8,29.7]	**-2.7**	[-7.2,1.7]
Greece	23.4	[18.8,28.1]	18.5	[14.5,22.5]	**4.9**	[-1.2,11.1]
Slovenia	22.8	[19.3,26.2]	20.2	[17.1,23.2]	**2.6**	[-2.0,7.2]
Poland	21.1	[12.4,29.7]	20.3	[18.1,22.4]	**0.8**	[-8.1,9.7]
N	30494		41753		72247	

Estimates based on average marginal effects after a logistic regression model, setting age and sex to the all-country/all-survey means.

Disability weights

As described in the main text / below, we create disability scales in two ways:

1. 'Latent disability' – based on how strongly each measure is associated with the other measures in the scale, using a two-parameter Item Response Theory (IRT) model;
2. 'Predicted disability' – based on how strongly each measure predicts self-reported general disability, in a logistic regression model.

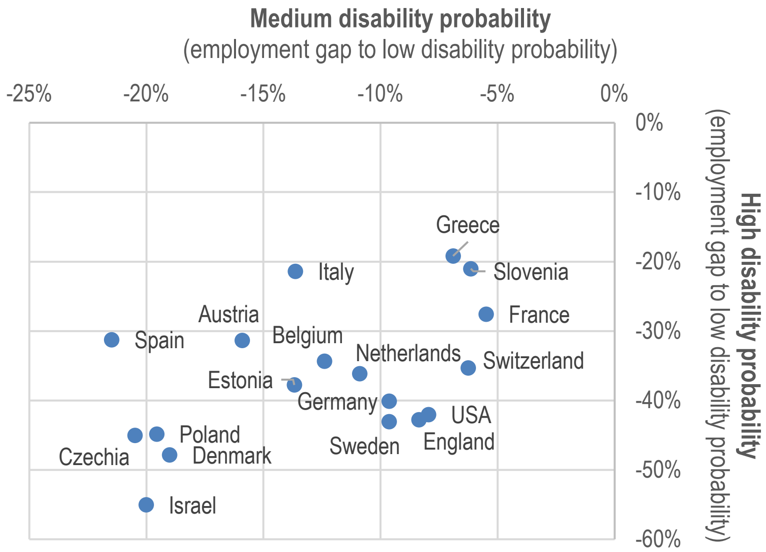
Table B2: Weights used in creating disability scales

The main text refers to the difference in how items are weighted. The details of this are shown in the table below; note that the scales that the weights are on are not comparable, but the respective strength of the weights for particular items within a given scale can be compared. So for example: 'mental ill-health' has a relatively strong weight in the predicted disability scale, whereas its discrimination parameter in the IRT model is relatively weak.

	Predicted disability		*Latent disability*			
	Weight	**p**	**Discrimination**	**p**	**Difficulty**	**p**
Mental ill-health (caseness)	0.705	[0.000]	1.01	0.00	1.43	0.00
Motor skills: walk 100m/one block	0.483	[0.000]	5.26	0.00	1.54	0.00
Motor skills: sit for 2hrs	0.500	[0.000]	2.12	0.00	1.55	0.00
Motor skills: get up from chair	0.650	[0.000]	2.59	0.00	1.08	0.00
Motor skills: climbing (partial)	0.793	[0.000]	2.69	0.00	0.81	0.00
Motor skills: climbing (severe)	0.955	[0.000]			1.58	0.00
Motor skills: stoop/kneel/crouch	0.817	[0.000]	2.55	0.00	0.76	0.00
Motor skills: reaching above shoulder	0.597	[0.000]	2.19	0.00	1.75	0.00
Motor skills: pulling/pushing large objects	0.650	[0.000]	3.02	0.00	1.35	0.00
Motor skills: lifting 5kg	0.988	[0.000]	2.71	0.00	1.20	0.00
Motor skills: picking up a small coin	0.121	[0.476]	2.02	0.00	2.46	0.00
Vision limitations (near and/or far sight)			0.56	0.00		
very good or excellent					ref	
good, very good or excellent	ref					
good					-0.74	0.00
fair	0.167	[0.003]			2.53	0.00
poor/impossible	0.458	[0.000]			4.92	0.00
Hearing limitations			0.44	0.00		
excellent					ref	
excellent, very good or fair	ref					
very good					-3.33	0.00
good					-0.13	0.00
fair	0.318	[0.000]			4.25	0.00
poor	0.550	[0.001]			8.80	0.00
Any ADL	0.667	[0.000]				
ADLs: Dressing			3.01	0.00	1.82	0.00
ADLs: Bathing or showering			4.10	0.00	1.98	0.00
ADLs: Eating			3.05	0.00	2.59	0.00
ADLs: Getting in or out of bed			3.39	0.00	2.03	0.00
ADLs: Using the toilet			3.64	0.00	2.26	0.00
ADLs: Walking limitations			5.51	0.00		
Can't walk 100m or a block					1.51	0.00
Can't walk across a room					2.13	0.00
1 IADL	0.775	[0.000]				
2+ IADLs	0.430	[0.324]				
IADLs: Preparing a hot meal			3.78	0.00	2.20	0.00
IADLs: Shopping for groceries			4.00	0.00	1.92	0.00
IADLs: Making telephone calls			2.55	0.00	2.80	0.00
IADLs: Taking medications			2.72	0.00	2.71	0.00
IADLs: Managing money			2.03	0.00	2.68	0.00

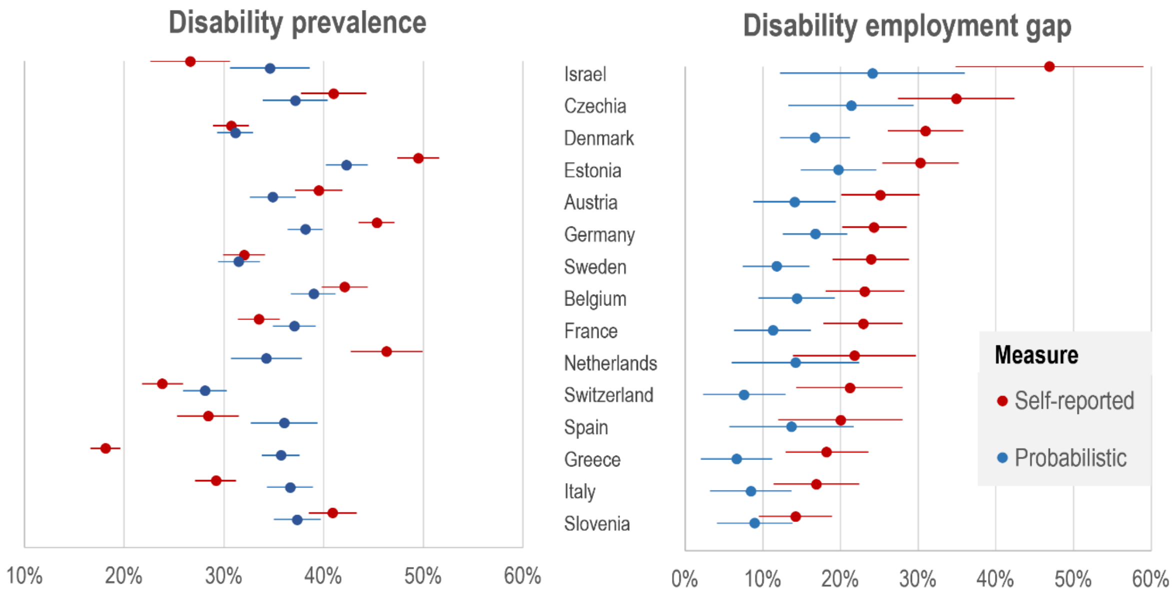
Source: ELSA-SHARE-HRS.

Figure B1: The difference between employment gaps of medium vs. high disability probabilities



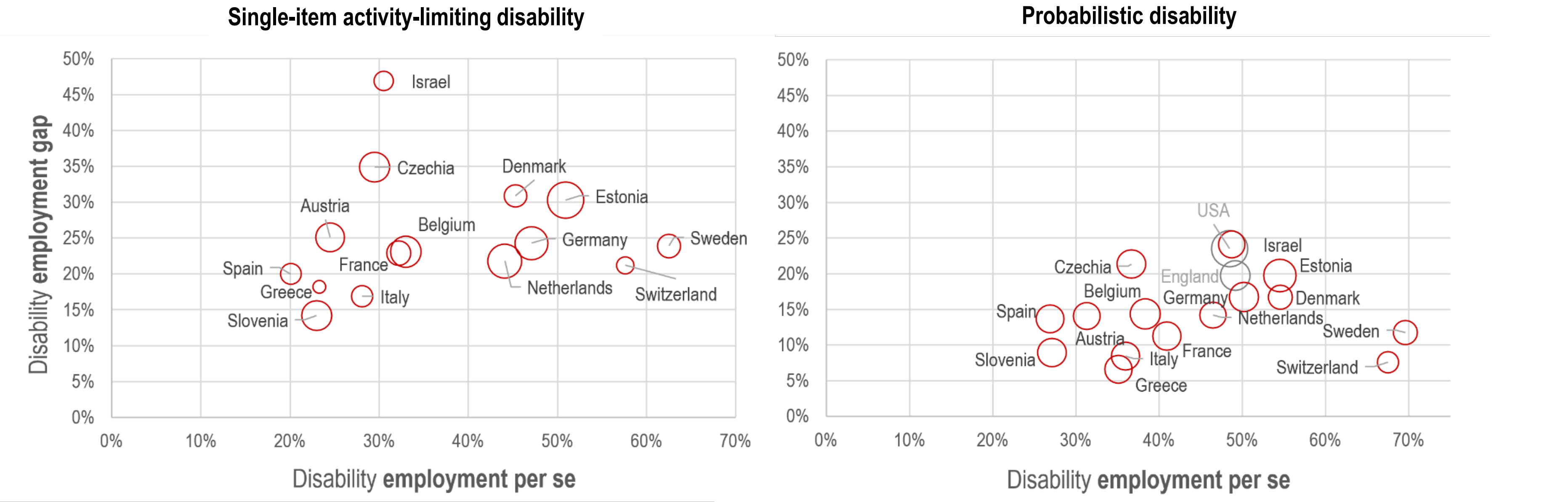
Disability probabilities are categorised as follows: 'low disability probability' = below the fixed threshold for the binary measure of predicted disability (32%); 'medium disability probability' = 32-70%; 'high disability probability' = 70-100%. Source: ELSA-SHARE. As for other figures, estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means.

Figure B2: The combined impact of methodological changes: the difference between using a probabilistic vs. single-item activity-limiting disability measure on prevalence and employment gaps among 50-69-year-olds



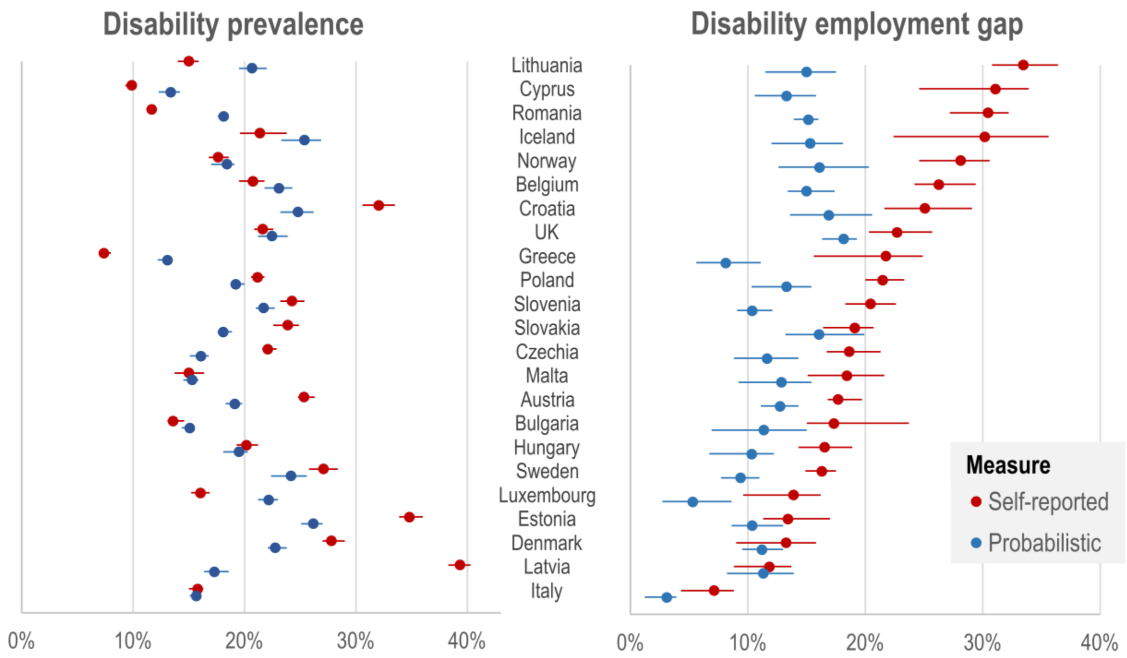
Source: Source: Authors' analysis of SHARE-ELSA-HRS data. Estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means. Probabilistic results are based on 1,000 replications (in which each person is considered 'disabled' or not according to their probability of reporting a disability).

Figure B3: How final disability employment results differ for single-item activity-limiting disability vs. probabilistic disability, for 50-69-year-olds



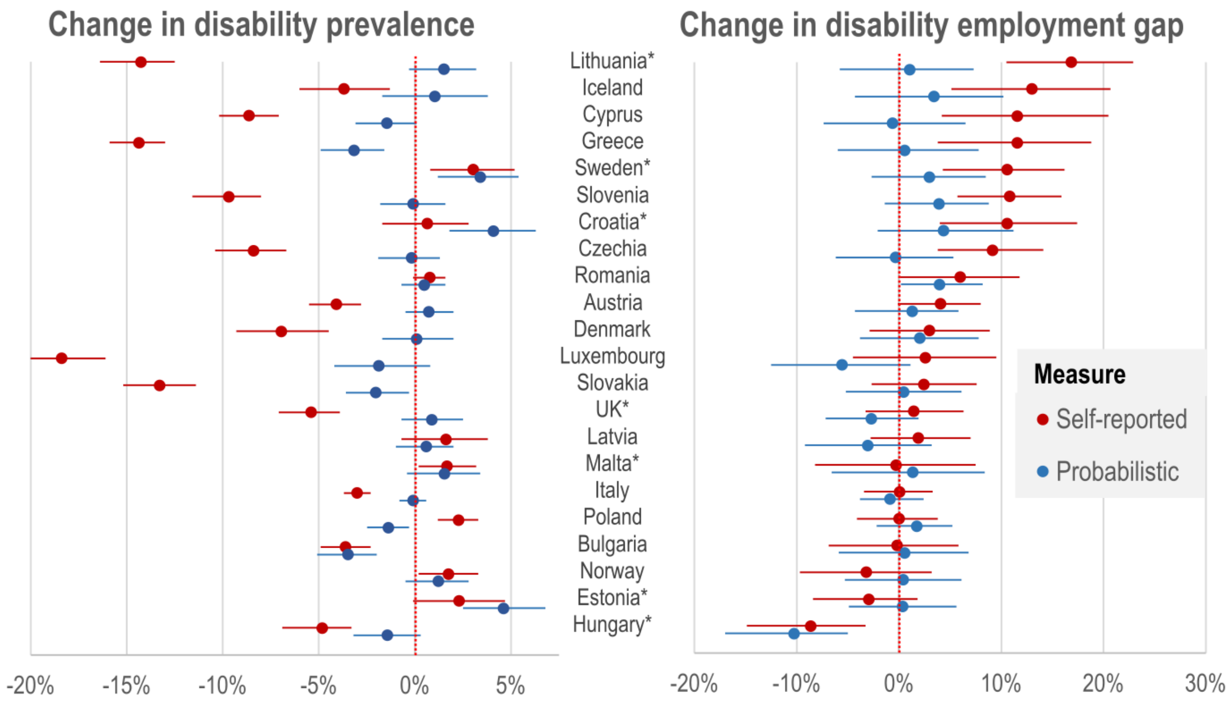
Source: Authors' analysis of SHARE-ELSA-HRS data. Bubble width represents the prevalence of disability (not area). Estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means. Probabilistic results are based on 1,000 replications (in which each person is considered 'disabled' or not according to their probability of reporting a disability).

Figure B4: How EHIS international comparisons change when using single-item activity-limiting disability vs. probabilistic disability based on predicted disability scales, for 20-69-year-olds in 2013-15



Source: EHIS. Estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means. Probabilistic results are based on 1,000 replications (in which each person is considered 'disabled' or not according to their probability of reporting a disability).

Figure B5: How trends in EHIS international comparisons change when using single-item activity-limiting disability vs. probabilistic disability based on predicted disability scales,
for 20-69-year-olds in 2018-20 vs. 2013-15



Source: EHIS. * indicates changes in the balance of survey mode between waves. Estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means. Probabilistic results are based on 1,000 replications (in which each person is considered 'disabled' or not according to their probability of reporting a disability).

Appendix B2: Exact values underlying figures in the main text

The following tables provide exact values underlying figures shown in the main text:

- Table B3: Table underlying Figure 2.1: The disability employment gap vs. the proportion of people prevented from working due to disability (PPWD), for 50-69 year olds in 2010-14
- Table B4: Table underlying Figure 3.1 (left panel): The prevalence of disability for 50-69 year olds in 2010-14
- Table B5: Table underlying Figure 3.1 (right panel): The disability employment gap for 50-69 year olds in 2010-14
- Table B6: Table underlying Figure 3.1 (not shown): The absolute employment rate for people with disabilities for 50-69 year olds in 2010-14
- Table B7: Table underlying Figure 3.4 (left panel): The disability prevalence rate for 50-69 year olds in 2010-14
- Table B8: Table underlying Figure 3.4 (right panel): The disability employment gap for 50-69 year olds in 2010-14
- Table B9: Table underlying Figure 3.4 (not shown): The absolute employment rate for people with disabilities for 50-69 year olds in 2010-14

Table B3: Table underlying Figure 1: The disability employment gap vs. the proportion of people prevented from working due to disability (PPWD), for 50-69 year olds in 2010-14

	Disability employment gap		PPWD	
Slovenia	14.2%	*[9.0 to 18.8]*	5.8%	*[3.7 to 7.7]*
Italy	16.9%	*[10.9 to 22.3]*	4.9%	*[3.2 to 6.6]*
Greece	18.2%	*[12.1 to 23.5]*	3.3%	*[2.2 to 4.3]*
Spain	20.0%	*[12.0 to 27.2]*	5.7%	*[3.2 to 7.9]*
Switzerland	21.2%	*[15.3 to 28.3]*	5.0%	*[3.5 to 6.9]*
Netherlands	21.8%	*[13.8 to 30.5]*	10.1%	*[6.4 to 13.9]*
France	22.9%	*[17.2 to 27.9]*	7.7%	*[5.9 to 9.5]*
Belgium	23.1%	*[18.6 to 28.4]*	9.7%	*[7.8 to 11.9]*
Sweden	23.9%	*[19.0 to 29.2]*	7.6%	*[6.0 to 9.5]*
Germany	24.3%	*[20.3 to 28.2]*	11.0%	*[9.1 to 12.9]*
Austria	25.1%	*[20.4 to 30.5]*	9.9%	*[7.9 to 12.1]*
Estonia	30.3%	*[25.6 to 35.7]*	15.0%	*[12.4 to 17.8]*
Denmark	30.9%	*[26.1 to 35.3]*	9.5%	*[7.8 to 11.1]*
Czechia	34.9%	*[26.9 to 43.3]*	14.3%	*[10.9 to 18.0]*
Israel	46.9%	*[32.8 to 57.3]*	12.5%	*[8.2 to 16.7]*

Estimates based on average marginal effects after a logistic regression model, setting age and sex to the all-country/all-survey means. Note that these estimates are bootstrapped (see main text/Appendix B2), so differ very slightly to the conventional standard errors shown in Table 1 and Table 2 of the main report.

Table B4: Table underlying Figure 3 (left panel): The prevalence of disability for 50-69 year olds in 2010-14

	Self-report		*Latent*		*Predicted*	
	Estimate	*95% CI*	*Estimate*	*95% CI*	*Estimate*	*95% CI*
Israel	27%	22.6,30.6	30%	26.2,34.3	28%	24.2,31.8
Czechia	41%	37.7,44.3	38%	35.2,41.5	37%	33.6,39.9
Denmark	31%	28.9,32.5	28%	26.5,30.1	27%	25.4,28.9
Estonia	50%	47.4,51.6	45%	42.8,47.0	45%	43.0,47.2
Austria	40%	37.1,41.9	35%	32.6,37.2	32%	29.7,34.1
Germany	45%	43.5,47.1	41%	39.2,42.8	41%	39.4,43.0
Sweden	32%	29.9,34.1	29%	26.8,30.7	30%	27.7,31.7
Belgium	42%	39.8,44.4	43%	40.8,45.4	43%	40.4,45.1
France	34%	31.4,35.6	36%	33.8,38.1	37%	34.6,39.0
Netherlands	46%	42.7,49.9	31%	27.5,34.7	32%	28.1,35.2
Switzerland	24%	21.7,25.8	23%	20.5,24.5	23%	20.8,24.9
Spain	28%	25.3,31.5	33%	29.5,36.0	34%	30.8,37.5
Greece	18%	16.6,19.6	40%	38.2,42.0	38%	36.3,40.1
Italy	29%	27.2,31.3	35%	32.6,37.1	38%	35.9,40.5
Slovenia	41%	38.5,43.3	37%	35.0,39.7	37%	34.8,39.5
USA			57%	55.7,58.1	53%	51.3,53.7
England			40%	38.1,41.7	39%	36.9,40.4

Source: ELSA-SHARE-HRS. Estimates based on average marginal effects after a logistic regression model, setting age and sex to the all-country/all-survey means.

Table B5: Table underlying Figure 3 (right panel): The disability employment gap for 50-69 year olds in 2010-14

	Self-report		***Latent***		***Predicted***	
	Estimate	*95% CI*	*Estimate*	*95% CI*	*Estimate*	*95% CI*
Israel	*47%*	*-59.0,-34.8*	*33%*	*-46.3,-19.7*	*41%*	*-52.0,-29.1*
Czechia	*35%*	*-42.4,-27.4*	*31%*	*-38.4,-23.5*	*32%*	*-39.6,-24.6*
Denmark	*31%*	*-35.8,-26.1*	*28%*	*-32.6,-23.4*	*30%*	*-34.3,-24.9*
Estonia	*30%*	*-35.2,-25.4*	*26%*	*-30.6,-21.2*	*26%*	*-30.6,-21.1*
Austria	*25%*	*-30.2,-20.1*	*24%*	*-28.8,-18.9*	*23%*	*-28.4,-18.4*
Germany	*24%*	*-28.5,-20.2*	*22%*	*-26.2,-18.2*	*22%*	*-26.3,-18.2*
Sweden	*24%*	*-28.8,-19.0*	*20%*	*-24.3,-15.0*	*19%*	*-23.4,-14.3*
Belgium	*23%*	*-28.2,-18.1*	*20%*	*-25.0,-15.3*	*22%*	*-26.4,-16.8*
France	*23%*	*-28.0,-17.8*	*16%*	*-20.5,-10.6*	*15%*	*-20.2,-10.4*
Netherlands	*22%*	*-29.7,-13.9*	*22%*	*-30.5,-13.9*	*22%*	*-29.8,-13.3*
Switzerland	*21%*	*-28.0,-14.3*	*13%*	*-19.4,-7.3*	*14%*	*-19.9,-7.7*
Spain	*20%*	*-28.0,-12.0*	*23%*	*-31.0,-15.7*	*26%*	*-33.1,-18.1*
Greece	*18%*	*-23.6,-12.9*	*8%*	*-12.8,-3.7*	*11%*	*-15.2,-6.1*
Italy	*17%*	*-22.4,-11.4*	*14%*	*-19.0,-8.4*	*17%*	*-21.7,-11.2*
Slovenia	*14%*	*-18.9,-9.5*	*12%*	*-17.1,-7.7*	*13%*	*-17.9,-8.5*
USA			*24%*	*-27.1,-21.7*	*27%*	*-29.4,-24.0*
England			*23%*	*-26.9,-19.1*	*26%*	*-29.6,-21.6*

Source: ELSA-SHARE-HRS. Estimates based on average marginal effects after a logistic regression model, setting age and sex to the all-country/all-survey means.

Table B6: Table underlying Figure 3 (not shown): The absolute employment rate for people with disabilities for 50-69 year olds in 2010-14

	Self-report		***Latent***		***Predicted***	
	Estimate	*95% CI*	*Estimate*	*95% CI*	*Estimate*	*95% CI*
Israel	*31%*	*19.4,41.6*	*42%*	*29.5,53.8*	*35%*	*25.2,45.4*
Czechia	*30%*	*23.1,36.0*	*31%*	*25.1,37.2*	*30%*	*23.8,36.1*
Denmark	*45%*	*40.9,49.7*	*46%*	*41.9,50.1*	*45%*	*40.2,48.8*
Estonia	*51%*	*46.8,55.0*	*52%*	*47.6,55.5*	*52%*	*47.9,55.7*
Austria	*25%*	*20.8,28.2*	*25%*	*21.3,29.0*	*25%*	*20.8,28.6*
Germany	*47%*	*43.7,50.6*	*48%*	*44.2,50.9*	*48%*	*44.2,50.9*
Sweden	*63%*	*57.7,67.2*	*64%*	*59.4,68.1*	*65%*	*60.2,68.7*
Belgium	*33%*	*29.2,36.9*	*36%*	*32.1,39.2*	*35%*	*31.0,38.6*
France	*32%*	*27.9,36.6*	*38%*	*34.1,42.3*	*39%*	*34.4,42.5*
Netherlands	*44%*	*37.5,50.7*	*41%*	*33.6,47.7*	*41%*	*34.3,48.1*
Switzerland	*58%*	*51.2,64.1*	*63%*	*56.9,68.3*	*62%*	*56.6,68.0*
Spain	*20%*	*14.1,26.0*	*20%*	*14.6,25.4*	*19%*	*13.6,24.3*
Greece	*23%*	*18.6,27.9*	*34%*	*30.9,37.8*	*33%*	*29.2,36.1*
Italy	*28%*	*23.7,32.5*	*32%*	*28.2,36.7*	*31%*	*27.0,35.2*
Slovenia	*23%*	*19.6,26.4*	*25%*	*21.5,28.5*	*24%*	*20.9,27.9*
USA			*50%*	*48.4,52.3*	*48%*	*46.1,50.2*
England			*47%*	*44.2,50.6*	*46%*	*42.2,48.8*

Source: ELSA-SHARE-HRS. Estimates based on average marginal effects after a logistic regression model, setting age and sex to the all-country/all-survey means.

Table B7: Table underlying Figure 6 (left panel): The disability prevalence rate for 50-69 year olds in 2010-14

	Fixed cut-off		***Probabilistic***	
	Estimate	*95% CI*	*Estimate*	*95% CI*
Israel	*28%*	*[24.2 to 31.8]*	*35%*	*[30.6 to 38.6]*
Czechia	*37%*	*[33.6 to 39.9]*	*37%*	*[33.9 to 40.4]*
Denmark	*27%*	*[25.4 to 28.9]*	*31%*	*[29.3 to 32.9]*
USA	*53%*	*[51.3 to 53.7]*	*47%*	*[45.9 to 48.3]*
Estonia	*45%*	*[43.0 to 47.2]*	*42%*	*[40.2 to 44.4]*
England	*39%*	*[36.9 to 40.4]*	*39%*	*[37.0 to 40.5]*
Spain	*34%*	*[30.8 to 37.5]*	*36%*	*[32.7 to 39.4]*
Austria	*32%*	*[29.7 to 34.1]*	*35%*	*[32.6 to 37.2]*
Germany	*41%*	*[39.4 to 43.0]*	*38%*	*[36.4 to 39.9]*
Belgium	*43%*	*[40.4 to 45.1]*	*39%*	*[36.7 to 41.2]*
Netherlands	*32%*	*[28.1 to 35.2]*	*34%*	*[30.7 to 37.8]*
Sweden	*30%*	*[27.7 to 31.7]*	*31%*	*[29.4 to 33.6]*
Italy	*38%*	*[35.9 to 40.5]*	*37%*	*[34.3 to 38.9]*
France	*37%*	*[34.6 to 39.0]*	*37%*	*[34.9 to 39.2]*
Switzerland	*23%*	*[20.8 to 24.9]*	*28%*	*[25.9 to 30.3]*
Slovenia			*37%*	*[35.0 to 39.7]*
Greece			*36%*	*[33.8 to 37.6]*

Source: ELSA-SHARE-HRS. Estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means. Probabilistic results are based on 1,000 replications (in which each person is considered 'disabled' or not according to their probability of reporting a disability).

Table B8: Table underlying Figure 6 (right panel): The disability employment gap for 50-69 year olds in 2010-14

	Fixed cut-off		*Probabilistic*	
	Estimate	*95% CI*	*Estimate*	*95% CI*
Israel	*41%*	*[29.1 to 52.0]*	*24%*	*[12.2 to 36.0]*
Czechia	*32%*	*[24.6 to 39.6]*	*21%*	*[13.3 to 29.4]*
Denmark	*30%*	*[24.9 to 34.3]*	*17%*	*[12.2 to 21.2]*
USA	*27%*	*[24.0 to 29.4]*	*24%*	*[20.8 to 26.3]*
Estonia	*26%*	*[21.1 to 30.6]*	*20%*	*[14.9 to 24.6]*
England	*26%*	*[21.6 to 29.6]*	*20%*	*[15.7 to 23.8]*
Spain	*26%*	*[18.1 to 33.1]*	*14%*	*[5.7 to 21.7]*
Austria	*23%*	*[18.4 to 28.4]*	*14%*	*[8.8 to 19.4]*
Germany	*22%*	*[18.2 to 26.3]*	*17%*	*[12.6 to 20.9]*
Belgium	*22%*	*[16.8 to 26.4]*	*14%*	*[9.4 to 19.3]*
Netherlands	*22%*	*[13.3 to 29.8]*	*14%*	*[6.0 to 22.4]*
Sweden	*19%*	*[14.3 to 23.4]*	*12%*	*[7.4 to 16.0]*
Italy	*16%*	*[11.2 to 21.7]*	*8%*	*[3.2 to 13.7]*
France	*15%*	*[10.4 to 20.2]*	*11%*	*[6.3 to 16.2]*
Switzerland	*14%*	*[7.7 to 19.9]*	*8%*	*[2.3 to 12.9]*
Slovenia			*9%*	*[4.1 to 13.8]*
Greece			*7%*	*[2.0 to 11.2]*

Source: ELSA-SHARE-HRS. Estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means. Probabilistic results are based on 1,000 replications (in which each person is considered 'disabled' or not according to their probability of reporting a disability).

Table B9: Table underlying Figure 6 (not shown): The absolute employment rate for people with disabilities for 50-69 year olds in 2010-14

	Fixed cut-off		*Probabilistic*	
	Estimate	*95% CI*	*Estimate*	*95% CI*
Israel	*35%*	*[25.2 to 45.4]*	*49%*	*[38.5 to 59.0]*
Czechia	*30%*	*[23.8 to 36.1]*	*37%*	*[29.9 to 43.4]*
Denmark	*45%*	*[40.2 to 48.8]*	*55%*	*[50.6 to 58.6]*
USA	*48%*	*[46.1 to 50.2]*	*48%*	*[46.3 to 50.6]*
Estonia	*52%*	*[47.9 to 55.7]*	*55%*	*[50.5 to 58.6]*
England	*45%*	*[42.2 to 48.8]*	*49%*	*[45.8 to 52.5]*
Spain	*19%*	*[13.6 to 24.3]*	*27%*	*[21.0 to 32.8]*
Austria	*25%*	*[20.8 to 28.6]*	*31%*	*[27.1 to 35.5]*
Germany	*48%*	*[44.2 to 50.9]*	*50%*	*[46.7 to 53.7]*
Belgium	*35%*	*[31.0 to 38.6]*	*38%*	*[34.5 to 42.2]*
Netherlands	*41%*	*[34.3 to 48.1]*	*46%*	*[39.6 to 53.3]*
Sweden	*64%*	*[60.2 to 68.7]*	*70%*	*[65.7 to 73.6]*
Italy	*31%*	*[27.0 to 35.2]*	*36%*	*[31.8 to 40.1]*
France	*38%*	*[34.4 to 42.5]*	*41%*	*[36.9 to 45.0]*
Switzerland	*62%*	*[56.6 to 68.0]*	*68%*	*[62.8 to 72.2]*
Slovenia	*24%*	*[20.9 to 27.9]*	*27%*	*[23.4 to 30.8]*
Greece	*33%*	*[29.2 to 36.1]*	*35%*	*[31.4 to 38.8]*

Source: ELSA-SHARE-HRS. Estimates based on average marginal effects after logistic regression models, setting age and sex to the all-country/all-survey means. Probabilistic results are based on 1,000 replications (in which each person is considered 'disabled' or not according to their probability of reporting a disability).

Appendix B3: A walkthrough of the differences in the disability employment rates between measures: the case of the Netherlands

In the main text, we find that the prevalence of disability among older people in the Netherlands changes substantially when using different methods – it is 45% when using single-item activity-limiting disability, but only 31% when using predicted disability. Despite this, though, the difference in the disability employment gap between these measures is negligible (a gap of 18 percentage points for single-item activity-limiting disability vs. 20 percentage points for predicted disability; the difference for disability employment rates is similarly small).

The formal algebraic expressions for the disability employment gap are less helpful than in Chapter 2, because they get messier. Instead, it is easier to understand by going through this step-by-step:

1. In the Netherlands, 14 percentage points fewer people have a 'predicted disability' than a single-item activity-limiting disability. Yet the overlap between measures is not perfect: not everyone with a predicted disability has a single-item activity-limiting disability. The 22% that only have a disability on the self-reported measure we term 'SR-only disability', and the 8% that only have a disability on the predicted-disability measure we term 'predicted-only disability'.

2. If we change to using ‘predicted disability’ rather than ‘single-item activity-limiting disability’, we are in effect…

   …moving people with SR-only disability from disability to non-disability, and…

   …moving people with predicted-only disability from non-disability to disability.

   The effect on the disability employment gap depends on both parts. Let us look at them in turn.

3. The effect of moving people with SR-only disability from disability to non-disability depends on how big this group is (22% of the population, compared to 46% consistently non-disabled and 23% consistently disabled people), combined with…

   …how their employment rate compares to consistently non-disabled people (i.e., those not disabled under both definitions). Their employment rate of 56% is lower than consistently non-disabled (which is 65%), so adding them to this group pulls the non-disabled employment rate down by 3%.

   …how their employment rate compares to consistently disabled people. Their employment rate of 56% is higher than that of consistently disabled people (which is 36%), so removing them from this group pulls the disabled employment rate down by 10%.

   In combination, this therefore increases the disability employment gap by 7 percentage points.

4. At the same time, we need to look at the effect of moving people with predicted-only disability from non-disability to disability. The 8% of people in this situation have an employment rate of 58%, so removing them from the non-disabled group increases the non-disabled employment rate by 1% and adding them to the disabled group increases the disability employment rate by 6%. This therefore reduces the disability employment gap by 5 percentage points.

5. Combining these, using ‘predicted disability’ rather than ‘single-item activity-limiting disability’ increases the disability employment gap by only 2 percentage points.

   There would be a greater change in the disability employment gap in the following situations:

   - If the employment rates among the SR-only and predicted-only disability groups were further away from 48%. If the employment rates in these groups was the same as consistently non-disabled people (65%), the gap would increase by 7 percentage points.
   - If there were different employment rates among the SR-only and predicted-only disability groups. The employment rates in these groups are very similar (indeed, the small difference in this case slightly offsets the greater numbers of people with SR-only disability). For example, if the employment rates for people with SR-only and predicted-only disability were 55% and 40% respectively, then changing to the predicted disability measure would increase the disability employment gap by 9 percentage points.

In summary: when changing disability measure, the extent of changes in the disability employment gap depends primarily on the employment rates among groups whose disability status changes (to a greater or lesser extent depending on how spread these groups are). It is therefore possible that large changes between measures in the prevalence of disability have little effect on the disability employment gap – but this is not a logical requirement, and noticeable changes in the disability employment gap from using different measures can be found (as we see from the EHIS results in Chapter 4).

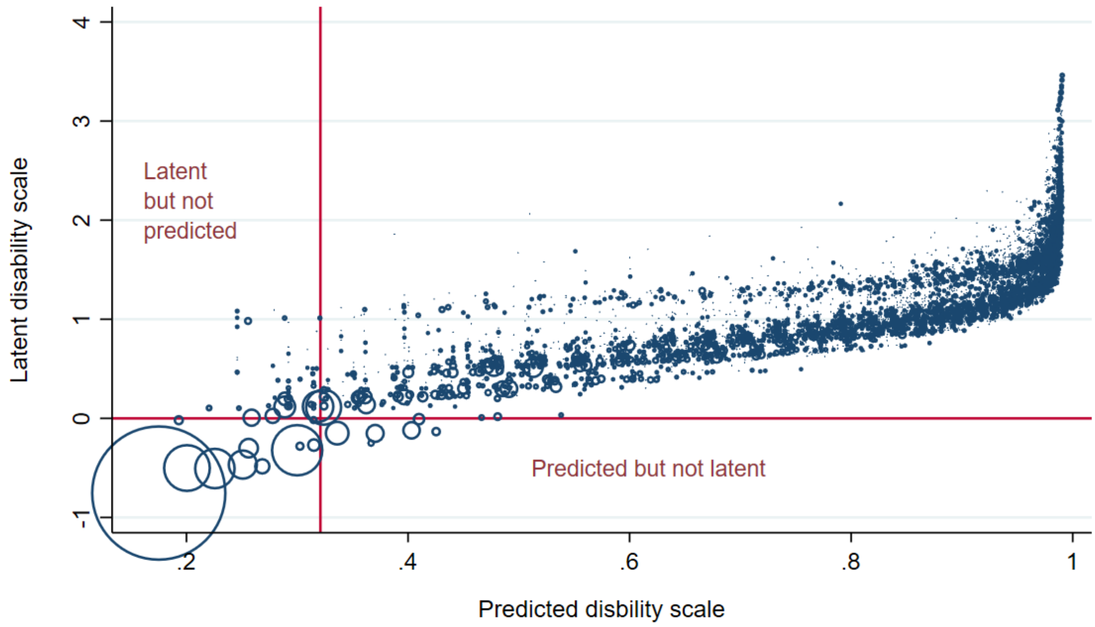
Appendix B4: The difference between the two multi-item scales

In the main text, we note: *"Within the two multi-item disability scales, there are also times when latent and predicted disability give different pictures – including the disability employment gaps in the USA, France, and Greece. In nearly all cases the disability employment gap for predicted disability is higher than it is for latent disability, likely to be because the disability weights better capture the genuine barriers people face. However, there are fewer differences between predicted disability and latent disability in terms of prevalence – only in Estonia and Czechia (and to a lesser extent the USA and England) is there a noticeable difference."*

In this appendix, we explore further why the two scales are so similar. It is not because the weights are the same: as shown in Appendix Table B2, we can see that the weights differ considerably between the two approaches, in exactly the ways we would expect. For example, mental health limitations have a very low weight for latent disability (lower than any motor skill or IADL), but a relatively high weight for predicted disability (higher than any ADL or most motor limitations). Picking up a small coin is the reverse, with a moderately high weight for latent disability, but a very low weight for predicted disability.

In Figure 3.2, we show how the two scales compare. Unsurprisingly there is a strong positive correlation between them; all indicators have positive weights, so the more impairments/activitity limitations that people report, the higher their disability scores on both scales. However, when turning these into binary disability categories, the differences in the scales rarely take people over the threshold to be counted as 'disabled'. 3% of older people are only over the threshold for predicted disability (the bottom-right corner of the figure); 4% are only over the threshold for latent disability (the top-left corner), but 37% are over the threshold on both scales. This is primarily because nearly everyone who reports any sort of impairment is categorised as 'disabled' for both measures. This is not automatic, but instead reflects the high self-reporting of overall disability compared to the prevalence of impairments/activitity limitations.

Figure 3.4. Comparing the predicted and latent disability scales



Source: Authors' analysis of SHARE-ELSA-HRS data. Note: the size of the bubbles reflects the number of people with this set of responses to the impairment questions.

In conclusion, the different multi-item scales produce different disability weights and scale scores, but similar binary disability variables (at least in this case) – and therefore produce similar comparisons across countries.

Appendix C: Other statistical details

Appendix C1: Full detail on bias in the disability employment gap and bias in the proportion prevented from working due to disability

Box 1 summarises the biases in two measures: the proportion prevented from working due to disability (PPWD) and the conventional disability employment gap. However, the full expressions for the bias are not given there as they do not reduce down to a simple expression, and they are therefore presented here in full.

As described in Box 1, the easiest way of examining the potential bias in different measures of the disability employment gap is to imagine three groups of people:

- Always-disabled people (who report a disability in all times/places),
- Never-disabled people (who never report a disability), and
- Sometimes-disabled people (who only sometimes report a disability).

We then look at how the disability employment gap changes if sometimes-disabled people <u>all</u> report a disability, vs. when <u>none</u> of them report a disability. As in Box 1, we refer to employment rates as E (so 'E_{SD}' refers to the employment rate among sometimes-disabled people), and prevalence rates as D (so 'D_D' refers to the prevalence rate of always-disabled people).

The change in the disability employment gap when reporting changes is as follows:

$$\Delta Gap = \frac{E_{PD}(P_{PD}P_D + P_{PD}P_{ND} + 2P_{PD}^2) - E_D(P_{ND}P_{PD} + P_{PD}^2) - E_{ND}(P_DP_{PD} + P_{PD}^2)}{(P_{PD} + P_{ND})(P_D + P_{PD})}$$

If the employment rate of sometimes-disabled people is the same as non-disabled people (E_{SD}=E_{ND}), then this simplifies slightly to:

$$\Delta Gap = (E_{PD} - E_D)\frac{P_{PD}}{(P_D + P_{PD})}$$

The change in the disability employment gap when reporting changes is as follows:

$$\Delta PPWD = \mathrm{P_{SD}}(E_{SD} - E_{ND})\frac{2P_{SD}P_D + P_D^2 + P_{SD}^2 + P_DP_{ND} + P_{SD}P_{ND}}{(P_{SD} + P_{ND})(P_{SD} + P_D)}$$

If the employment rate of sometimes-disabled people is the same as non-disabled people (E_{SD}=E_{ND}), then as Box 1 says, this expression is zero – which is clear from the expression above.

Appendix D: Data details

Appendix D1: Further details on SHARE-ELSA-HRS

Appendix A6 expains how to access the SHARE-ELSA-HRS data, and the RAND harmonised versions of the three datasets. These surveys are based on the same model and share many features of sample design and many identical questions. Further details on e.g. sampling can be found in their cohort profiles (Börsch-Supan *et al.*, 2013; Sonnega *et al.*, 2014; Steptoe *et al.*, 2013). We use data for 2010-14; the year varies depending on the exact timing of the wave that has the full set of data that we require. (It is 2010 for the USA, 2014 for England, 2015 for Greece, Belgium, Poland, Slovenia and Estonia, and 2013 for Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Switzerland, Israel, and Czechia).

In terms of the measures used, we found it necessary to go beyond the RAND harmonised versions of the data – partly to get a wider range of variables, and partly because there are issues of comparability for some variables in the RAND harmonised dataset:

Employment status: The greatest problems for comparability between SHARE-ELSA-HRS is employment status, which is not consistent between surveys:

- In ELSA/HRS, respondents were asked if they were working ‘at the present time’ (HRS) or in the ‘last month’ (ELSA);
- In SHARE, follow-up respondents (who had responded in a previous wave) were asked if they had worked since the last interview, usually two years ago, and if so they are asked for the exact timings of moves in/out of work.

We therefore treat SHARE respondents as out-of-work if they stopped work longer ago than the previous calendar month, which reduces the employment rate of those affected by 4 percentage points.

Single-item activity-limiting disability: the disability measure in SHARE is the GALI question: “For the past six months at least, to what extent have you been limited because of a health problem in activities people usually do?”, where respondents could choose between “Severely limited”, “Limited, but not severely”, or “Not limited”. (Note that ELSA and HRS do not include this question – despite an unconvincing attempt to use a work disability question in the RAND Harmonized data – so they are not considered in Chapter 2 of the main report).

Within this GALI question, we have focused on those who describe any limitation – that is, ‘limited, but not severely’ as well as ‘severely limited’. This is primarily because it captures the wider group of people with disabilities that is most often of interest when people look at the disability employment gap (it also leads to slightly more precise estimates, because the sample size of people with disabilities is larger). Users should however also consider the extent to which countries do better/worse for those with more severe impairments, as discussed in Chapter 3.

Other impairment/activitity limitation/symptom measures: When constructing disability scales in Chapter 3, we use a series of specific impairments/activitity limitations/symptom measures:

- 6 measures of Activities of Daily Living (ADLs) – which were described in an early ELSA report (Breeze and Lang, 2008) as follows: *“The original scale of ADLs was developed by Katz and colleagues (Katz et al., 1963) who described them as ‘activities which people perform habitually and universally’ (p. 94). The activities covered in ELSA are: dressing, including putting on shoes and socks; walking across a room; bathing or showering; eating, such as cutting up food; getting in or out of bed; and using the toilet, including getting up or down.”*;
- 6 measures of Instrumental Activities of Daily Living (IADLs)) – which were described in an early ELSA report (Breeze and Lang, 2008) as follows: *"IADLs are everyday tasks involving a mix of cognitive and physical competences. The list used in ELSA comes from one developed and validated by Lawton and Brody (1969) to reflect what they termed ‘instrumental selfcare’. They are: preparing a hot meal; shopping for groceries; making telephone calls; taking medications; doing work around the house or garden; and managing money, such as paying bills or keeping track of expenses. An additional activity introduced into the US Health and Retirement Survey referred to using a map to figure out how to get around in a strange place (Fonda and Herzog, 2004); this activity has not been used in this chapter because it did not group consistently with the other IADLs."* Mirroring the ELSA team’s view, we have also excluded the ‘using a map’ measure*;*
- 10 motor skill impairments) – which were described in an early ELSA report (Breeze and Lang, 2008) as follows: *"Problems with motor skills and strength may be potential precursors to restrictions on participation. Respondents in ELSA are asked about ten items referring 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. The ten items are: walking 100 yards; getting up from a chair after sitting for long periods; climbing several flights of stairs without resting; climbing one flight of stairs without resting; stooping, kneeling or crouching; pulling or pushing large objects like a living-room chair; lifting or carrying weights over 10 pounds, like a heavy bag of groceries; reaching or extending arms above shoulder level; sitting for about two hours; and picking up a small coin from a table";*
- Single-item summary measures of each of vision limitations (a single derived variable combining the variables rhefrnd and rhepap) and hearing limitations (a grouped version of rehear); and
- A mental health scale. The mental health scales are slightly different between surveys (CES-D in HRS/ELSA vs. Euro-D in SHARE), but both are well-validated scales for depressive symptoms, and the fact that the scales are different is unlikely to affect our results.[4]

For further details about which items were used and how they operationalised, please see the replication file detailed in Appendix A6, and particularly the file ‘2_cleaningXsurvey_labelling.do’, which takes the raw variables from the SHARE-ELSA-HRS files and prepares them for analysis. (Note however that not all of the variables that are created here are used in analysis, as for some variables the comparability between surveys was felt to be questionable).

Appendix D2: Further details on EHIS

Appendix A6 expains how to access the EHIS data from Eurostat. As we said in the main report (Chapter 2), there are major concerns about the comparability of EHIS data between countries – it is much less

[4] While mental health has been used in previous HRS-SHARE comparisons (Jadhav and Weir, 2017; Riumallo-Herl *et al.*, 2014), the comparability of these mental health measures between surveys is limited (Courtin *et al.*, 2015). However, in our case, we find that the overall prevalence of disability is higher in ELSA/ HRS vs. SHARE, but this cannot be explained by the mental health scale (which has similar prevalence in ELSA/HRS vs. SHARE).

comparable than SHARE (with different survey modes, different sampling approaches, and sometimes even different question wording, all of which will influence the results).

In terms of the measures used, readers may want to note:

- *Self-reported general disability:* the disability measure in EHIS is the same GALI measure used in SHARE, although because EHIS is underpinned by an EU Regulation, there are likely to be some small variations in wording between countries (for example, the EHIS wave 2 documentation suggests minor variations in Bulgaria, France, Netherlands, Norway). As for SHARE, we have focused on those who describe any limitation – that is, 'limited, but not severely' as well as 'severely limited'.
- *Impairments/activity limitations:* fewer measures are available in EHIS compared to SHARE, and we are therefore restricted to using:
    - A mental health scale (PHQ, grouped into Minimal depression/Mild depression/Moderate depression/Moderately severe depression/Severe depression),
    - A measure of difficulty in walking (from combining variables pl6 and pl7, which refer to difficulty walking on level ground and up steps),
    - Difficulty seeing (variable pl2) and hearing (from combining the variables pl4 and pl5), and
    - Intensity of bodily pain in past 4wks (variable pn1).

Again, for further details about which items were used and how they operationalised, please see the replication file detailed in Appendix A6.

Appendix D3: UK disability and employment data

In Chapter 3 of the main report, we present a case study of official figures on disability and employment in the UK. These figures are taken from the UK Labour Force Survey (LFS), as reported in the Office of National Statistics series 'A08: Labour market status of disabled people'[5].

Note that there are discontinuities in the LFS data (shown by different line styles in the graph in the main text). After each discontinuity the new series shows a different prevalence of disability than before, which visually makes it difficult to see the increase in disability over the entire period. In the figures in the main text we have therefore converted this into a single 'chained' series, where the end point of the earlier trend is set to the same value as the most recent point of the more recent trend. So for example:

- The prevalence of disability in Jan-Mar 2017 and Apr-Jun 2017 was 17.2% and 17.3% respectively, but there was then a discontinuity in the data, after which the prevalence in Jul-Sep 2017 was 18.2%.
- We assume that there are no changes in disability across the discontinuity, and in the chained series set the Apr-Jun 2017 value to 18.2%, and continue the trends backwards from this point, so that Jan-Mar 2017 is 18.1% (0.1 percentage points lower than Apr-Jun 2017) etc.

This deals with the discontinuities in the data, while also allowing us to more easily visually see the trends over this period. The exact figures in each year can be found in the Office of National Statistics public tables.

5
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/labourmarketstatusofdisabledpeoplea08

End matter for appendices

References

Börsch-Supan, A., et al. (2013), 'Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)', *International Journal of Epidemiology*, 42: 4, 992-1001.

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

Breeze, E. and Lang, I. (2008), 'Physical functioning in a community context', in J. Banks and et al (eds.), *Living in the 21st Century: older people in England (The 2006 English Longitudinal Study of Ageing, Wave 3)*, London: Institute for Fiscal Studies.

Courtin, E., et al. (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: 4, 287-304.

Efron, B. and Tibshirani, R.J. (1993), *An introduction to the bootstrap*, London: Chapman and Hall.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The Elements of Statistical Learning [2nd edition]*: Springer.

Hesterberg, T. (2014), What Teachers Should Know about the Bootstrap: Resampling in the Undergraduate Statistics Curriculum, *arXiv*. https://arxiv.org/abs/1411.5279

Jadhav, A. and Weir, D. (2017), 'Widowhood and Depression in a Cross-National Perspective: Evidence from the United States, Europe, Korea, and China', *Journals of Gerontology Series B - Psychological Sciences and Social Sciences*.

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

Kolenikov, S. (2010), 'Resampling Variance Estimation for Complex Survey Data', *The Stata Journal*, 10: 2, 165-199.

Lee, N., Cadogan, J.W. and Chamberlain, L. (2013), 'The MIMIC model and formative variables: problems and solutions', *AMS Review*, 3: 1, 3-17.

Riumallo-Herl, C., et al. (2014), 'Job loss, wealth and depression during the Great Recession in the USA and Europe', *International Journal of Epidemiology*, 43: 5, 1508-1517.

Sonnega, A., et al. (2014), 'Cohort Profile: the Health and Retirement Study (HRS)', *International Journal of Epidemiology*, 43: 2, 576-585.

Steptoe, A., et al. (2013), 'Cohort Profile: The English Longitudinal Study of Ageing', *International Journal of Epidemiology*, 42: 6, 1640-1648.