

Cross-Country Differentials in Work Disability Reporting Among Older Europeans

Viola Angelini · Danilo Cavapozzi · Omar Paccagnella

Received: 26 May 2011 / Accepted: 26 May 2011 / Published online: 15 June 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Descriptive evidence shows that there is large cross-country variation in self-reported work disability rates of the elderly in Europe. In this paper we analyse whether these differences are genuine or they just reflect heterogeneity in reporting styles. To shed light on the determinants of work-disability differentials across countries, we combine a wide set of individuals' socioeconomic and health status characteristics with macro-economic indicators describing the institutional background of the country of residence.

Keywords Work disability · Vignettes · Reporting heterogeneity

1 Introduction

Demographic ageing will be a major determinant of long run economic development in Europe. The extent of the demographic changes, which include a combination of increasing life expectancy and low birth rates, is dramatic and will deeply affect future labour and financial markets. The expected strain on public budgets and especially social security has already received prominent attention and both the political debate and the scientific literature aim at finding sustainable reforms to solve the financial unbalance of

V. Angelini (✉)

Department of Economics, Econometrics and Finance, University of Groningen, Groningen,
The Netherlands
e-mail: v.angelini@rug.nl

V. Angelini

Netspar, Tilburg, The Netherlands

D. Cavapozzi

Department of Economics and Management, University of Padua, Padua, Italy
e-mail: danilo.cavapozzi@unipd.it

O. Paccagnella

Department of Statistical Sciences, University of Padua, Padua, Italy
e-mail: omar.paccagnella@unipd.it

public pensions (Economic Policy Committee 2000; Gruber and Wise, 2004; Bonoli and Shinkawa 2005). The design of employment policies targeted to persons with disabilities plays an important role when pursuing the financial equilibrium of modern welfare states. Since the normal retirement age is gradually increasing throughout Europe, disability schemes may constitute an alternative path for early exits from the labour force. Indeed, enrollment in disability programs often entails the permanent receipt of benefits and may act as a bridge towards the eligibility to classical labour retirement schemes.

As Burkhauser and Daly (2002) emphasize, the onset of a disability does not automatically imply the inability of carrying out a job for most people and work-based policies can significantly increase employment. However, the design of efficient reforms in this area requires a better understanding of the determinants of work limiting health problems among older workers. To this aim researchers have started to analyse data from general surveys, where individuals are asked to self-report their work disability status, together with their health and socio-economic characteristics. While physical measures of health are often related to specific features of a given health domain and provide information on health rather than work capacity, the advantage of using this self-reported measure is that it summarizes in a single index a variety of factors determining the work disability status of individuals. Indeed, as pointed out by Haveman and Wolfe (2000), a motor disability may turn out to be less limiting in the labour market activities than much milder health limitations, such as back pain, combined with low education levels. At the same time, self-assessments of work disability are highly subjective and this implies that, if different individuals have different beliefs about the concept of disability, their self-reports will result to be incomparable. This heterogeneity in reporting styles is called differential item functioning (DIF). Several papers (Kapteyn et al. 2007; King et al. 2004; King and Wand 2007) show how vignettes can be used to control for DIF and assess whether differences in disability rates across countries and socio-economic groups are genuine or they just reflect differences in response scales. For example, Kapteyn et al. (2007) show that allowing for heterogeneity in reporting styles leads to reduce by one half the raw differential in the disability rates between the Netherlands and the US.¹

In this paper we study the determinants of work disability reporting using data from the 2004 wave of the SHARE and anchoring vignettes on eight European countries, collected under the COMPARE project. SHARE is the first European multi-country dataset containing data on vignettes and allows us to investigate whether the differences in reporting styles found by Kapteyn et al. (2007) between the US and the Netherlands are present even when we compare relatively similar countries in continental Europe. Börsch-Supan (2005) documents that continental Europe is characterized by large cross-country heterogeneity in disability insurance enrolment rates, which cannot be explained by differences in the health status or in the demographic composition of the population. Our analysis intends to shed light on these differentials in work-disability rates across European countries by combining a wide set of individual socioeconomic and health status characteristics with macroeconomic indicators describing the institutional background of the country of residence.

The plan of the paper is as follows. Section 2 describes the data used in the analysis and reports some descriptive statistics. Section 3 presents the Hopit model. Section 4 presents and discusses the main estimation results as well as a counterfactual simulation. Section 5 concludes the paper and draws some comments on basis of our findings.

¹ The vignette methodology is also employed in Van Soest et al. (2006), who explicitly control for the presence of unobserved heterogeneity in the process determining individual response styles.

2 Data and Descriptive Statistics

Data are drawn from the first wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) on individuals aged 50 or over in eight European countries (Sweden, Germany, The Netherlands, Belgium, France, Spain, Greece and Italy). SHARE collects detailed information about a number of aspects that characterize the socioeconomic condition of the elderly, such as physical and mental health, employment, wealth and social support (see Börsch-Supan et al. 2005, 2008). As part of the COMPARE project, after the personal interview (CAPI, Computer Assisted Personal Interview), a subset of respondents are asked to fill an additional paper and pencil questionnaire, which contains questions on their own disability status and that of hypothetical persons described in particular situations and conditions (the *anchoring vignettes*). The vignettes cover three different domains of work disability, namely pain, affect and heart diseases, and were collected both in the first and in the second wave of the survey. However, while the 2004 questionnaire included three vignettes for each domain of work disability, in 2006 this number was reduced to just one. In this paper we will use data from the first wave in order to exploit the information gain resulting from a wider set of vignettes when correcting for heterogeneity in reporting styles.

In the survey, respondents are first asked to report the severity of any work limiting health problems they might have; the exact wording of the question is: “Do you have any impairment or health problem that limits the amount or kind of work you can do?”. Then, they are asked to evaluate the extent to which they think that the people described in the nine vignettes are limited in the kind or amount of work they can do.

For each domain of work disability, three vignettes are presented. Examples are:

- Pain: “[Kevin] suffers from back pain that causes stiffness in his back especially at work but is relieved with low doses of medication. He does not have any pains other than this generalized discomfort.”
- Affect: “[Anthony] generally enjoys his work. He gets depressed every 3 weeks for a day or two and loses interest in what he usually enjoys but is able to carry on with his day-to-day activities on the job.”
- Heart disease: “[Eve] has had heart problems in the past and she has been told to watch her cholesterol level. Sometimes if she feels stressed at work she feels pain in her chest and occasionally in her arms.”

How much is [Kevin/Anthony/Eve] limited in the kind or amount of work [he/she] could do?

Both for the self-evaluation and the vignettes the possible answers are “*none*”, “*mild*”, “*moderate*”, “*severe*” and “*extreme*”.

For identification, we keep only those respondents who give a valid answer to both the self-evaluation question and at least one vignette. We also select respondents up to 64 years older, so our final sample includes 2,464 individuals. Table 1 shows that our respondents are prevalently females (about 56%), they tend to live with a partner (83%) and they are on average 56.4 years old. More than 50% of them are either overweight or obese and the percentage of people who report having limitations with instrumental activities of daily living is only 7.5.

Figure 1 shows that there is a large cross-country variability in self-reported work disability rates, where we define a person as disabled if she reports to be moderately, severely or extremely limited in the kind or amount of work she can do. In particular, while the percentage of respondents who report to have work limiting health problems is over

Table 1 Description of the variables included in the regressions

Variable	Description	Mean	SD
Country of residence			
SE	Dummy = 1 if the person lives in Sweden	0.093	0.290
NL	Dummy = 1 if the person lives in The Netherlands	0.128	0.334
BE	Dummy = 1 if the person lives in Belgium	0.123	0.328
DE	Dummy = 1 if the person lives in Germany (baseline)		
FR	Dummy = 1 if the person lives in France	0.178	0.382
ES	Dummy = 1 if the person lives in Spain	0.098	0.297
IT	Dummy = 1 if the person lives in Italy	0.099	0.298
GR	Dummy = 1 if the person lives in Greece	0.174	0.379
Demographics			
Male	Dummy = 1 if the person is male	0.438	0.496
Age	(Age-56)/10	0.040	0.468
Age2	[(Age-56)/10] ²	0.221	0.314
Education			
Low_educ	Dummy = 1 if the ISCED code is at most 2 (baseline)		
Med_educ	Dummy = 1 if the ISCED code is 3	0.300	0.458
High_educ	Dummy = 1 if the ISCED code is at least 4	0.245	0.430
Employment			
Ever_worked	Dummy = 1 if the person has ever worked in her life	0.917	0.276
Civil status			
Partner	Dummy = 1 if the person has a cohabiting partner	0.828	0.378
Physical and mental health			
Overweight	Dummy = 1 if the person is overweight	0.431	0.495
Obese	Dummy = 1 if the person is obese	0.171	0.376
Smoking	Dummy = 1 if the person currently smokes	0.262	0.440
Arthritis	Dummy = 1 if the person suffers from arthritis, osteoarthritis or rheumatism	0.153	0.360
Chronic	Dummy = 1 if the person has at least two chronic diseases	0.286	0.452
Symptoms	Dummy = 1 if the person has at least two symptoms	0.290	0.454
Mobility	Dummy = 1 if the person has at least two mobility limitations	0.365	0.482
ADL	Dummy = 1 if the person has at least one limitation with activities of daily living	0.044	0.206
IADL	Dummy = 1 if the person has at least one limitation with instrumental ADL	0.075	0.264
Eurodcat	Dummy = 1 if the person is depressed according to the EURO-D scale	0.224	0.417
Fluency and handgrip tests			
Fluency	Number of animals cited in the fluency test (divided by 10) ^a	1.998	0.706
Recall	Number of words recalled in the ten words recall test	5.306	1.646
Grip	Measure of handgrip strength (divided by 10) ^b	3.671	1.219
Macroeconomic indicators (Source: Eurostat)			
Ind_prices	Harmonized index of consumer prices	2.161	0.660
Emp_5564	Employment rate for individuals aged 55-64	0.410	0.102
Dis_funct	Public expenditure for disability function as % of GDP	2.214	1.210

^a In the fluency test respondents are asked to name as many animals as they can think of in one minute

^b Two measurements of handgrip strength are taken for each hand and recorded according to a scale spanning from 0 to 100. We take the highest measurement

25% in Sweden, Germany and Spain, it is around 12% in the Netherlands and 6% in Greece. These cross-country differences can be only partly explained by institutional factors such as enrolment and eligibility rules, as suggested by Börsch-Supan (2008). The variation is very large, especially given that we are considering relatively similar countries in continental Europe, and they raise some doubts about the inter-cultural comparability of self-reported measures of work disability: is the variation observed in the data genuine or does it just reflect differences in reporting styles? A *prima facie* evidence in favour of the idea that reporting heterogeneity might play a role is provided by Fig. 2. The figure shows the evaluation that respondents in different countries give of the hypothetical persons described in three of the vignettes (Kevin, Anthony and Eve). It is interesting to note that Sweden and Spain, where we observe the highest work disability rates, are also the countries with the highest proportions of respondents evaluating the persons described in the vignettes as work disabled. In all countries “Eve” is considered as the one with the most severe work limiting health problems, while “Anthony” is rated as the one less work disabled (only in Sweden there is no statistical difference between how Kevin and Anthony are evaluated).

3 The Hopit Model

We use the econometric specification introduced by King et al. (2004) in its parametric version, the so-called hierarchical ordered probit model (Hopit). The model consists of two components: a self-assessment equation and a vignette equation.

3.1 Self-Assessment Component

Let us denote with Y_i^* the level of work disability perceived by individual $i = 1, \dots, n$:

$$\begin{aligned} Y_i^* &= X_i\beta + \varepsilon_i; \\ \varepsilon_i|X_i &\sim N(0, 1), \end{aligned} \tag{1}$$

where X_i is a vector of observable individual characteristics and ε_i is a normally distributed error term encompassing unobserved factors relevant for the determination of work disability levels. In the data we do not observe the latent variable Y_i^* but an ordered categorical variable Y_i , which takes values 1 (“none”), 2 (“mild”), 3 (“moderate”), 4 (“severe”) and 5 (“extreme”) through the mechanism:

$$Y_i = j \quad \text{if } \tau_i^{j-1} < Y_i^* \leq \tau_i^j, \quad j = 1, \dots, 5. \tag{2}$$

The main difference between the standard ordered probit and the Hopit model is that the thresholds τ_i^j are individual-specific and reflect individual characteristics:

$$\begin{aligned} \tau_i^0 &= -\infty; \quad \tau_i^5 = \infty; \\ \tau_i^1 &= X_i\gamma^1; \end{aligned} \tag{3}$$

$$\tau_i^j = \tau_i^{j-1} + \exp(X_i\gamma^j), \quad j = 2, 3, 4. \tag{4}$$

3.2 Vignette Component

To separately identify the parameters in β and γ , we need to use the additional information provided by vignette evaluations. As mentioned above, each respondent in our sample is

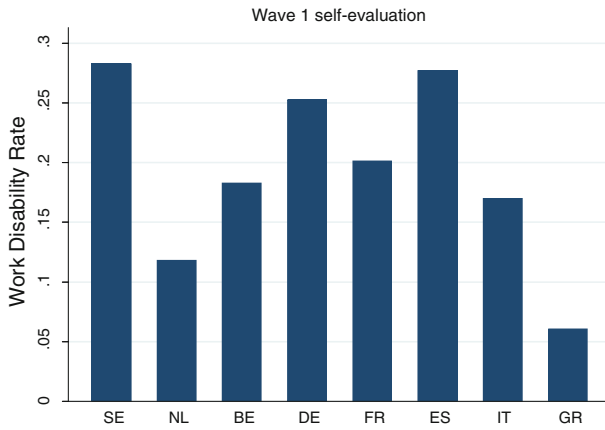


Fig. 1 Self-reported work disability rates by country

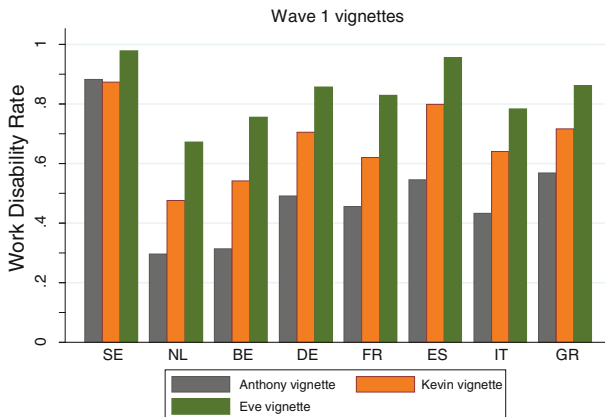


Fig. 2 Proportions of respondents who rate Anthony, Kevin and Eve as work disabled

asked to answer nine vignette questions describing hypothetical persons with different levels of work disability. Let us denote with Z_{il}^* , $l = 1, \dots, 9$ the underlying work disability of the person described in the vignette l as perceived by respondent i . We assume that:

$$\begin{aligned} Z_{il}^* &= \theta_l + v_{il}; \\ v_{il} &\sim N(0, \sigma_v^2), \end{aligned} \tag{5}$$

where v_{il} is a normally distributed error term independent of ε_i . The assumption of *vignette equivalence* implies that the situation described in the vignettes is, on average, perceived by respondents in the same way and formally restricts θ_l not to vary over i . As a result, the variability in the vignette evaluations provided by respondents is only due to the error term v_{il} in Eq. (5) and to the heterogeneous thresholds defined in Eqs. (3) and (4). Even in this case, Z_{il}^* is unobservable and we observe instead vignette ratings on a scale that goes from 1 (“none”) to 5 (“extreme”). Z_{il}^* is related to Z_{il} as follows:

$$Z_{il} = j \quad \text{if } \tau_i^{j-1} < Z_{il}^* \leq \tau_i^j, \quad j = 1, \dots, 5. \quad (6)$$

The two components of the model are connected via the utilization of the same set of thresholds in (2) and in (6). By restricting the thresholds to be the same, we assume that the same reporting styles are used both for self-assessments and vignette evaluations. This hypothesis is commonly referred to as the *response consistency* assumption.

There is a growing literature discussing the validity of the response consistency assumption (see Van Soest et al. 2007; Bago d'Uva et al. 2009). Van Soest et al. (2007) analyze drinking behaviour and extend the standard Hopit model by relaxing the response consistency assumption. They formally allow for the fact that reporting styles used by individuals to evaluate themselves may be different from those adopted to evaluate vignettes. The parameter identification relies on the availability of an objective indicator of the domain of interest, which is combined with self-assessments and vignette evaluations. An objective indicator may be qualified as suitable in this framework when it is unaffected by reporting heterogeneity and it is driven by the same underlying latent process determining self-assessments (one-factor assumption). To this end, Van Soest et al. (2007) take advantage of the information on the number of drinks consumed by the respondents and find that vignette based corrections appear quite effectively in bringing objective and subjective measures closer together. In the context of drinking behavior, which is clearly unidimensional, assuming that the proposed indicator is objective and that the one-factor assumption holds is plausible. However, work-disability is a multi-dimensional concept influenced by a combination of demographic, health and labour demand characteristics: thus, it is hard to think about an objective indicator satisfying the one-factor assumption and making it possible to test response consistency formally in our sample. The objective indicator used by Datta Gupta et al. (2010), namely grip strength, is clearly an unsatisfactory proxy.

Following King et al. (2004), the parameters of the self-assessment and the vignette components are jointly estimated by conditional maximum likelihood. The estimation is carried out using the gllamm (Generalized Linear Latent and Mixed Models) program of the STATA software (Rabe-Hesketh et al. 2004; Rabe-Hesketh and Skrondal 2008).

3.3 Counterfactuals

The main objective of the vignette approach is to estimate the DIF of each respondent and correct for it. Once model estimates are available, corrections for DIF are straightforward. The researcher can define a benchmark (for instance, a country) according to the explanatory variables included in the model and then compute adjusted distributions of the observed variable for all respondents, using the benchmark scale instead of respondents' own scales. The benchmark scale is calculated by predicting the thresholds under the assumption that all respondents belong to the benchmark group, regardless of their actual condition. Let us suppose that a dummy variable identifies the benchmark group of interest. The benchmark scale is defined by setting such dummy to 1 for all respondents when predicting the thresholds.² As for the other variables, we consider their actual values in the sample. The adjusted distributions of the self-evaluation reports are now defined according to a common scale and they can be compared meaningfully since differences across groups

² A similar approach can be used when the benchmark group is defined by a specific value of a variable treated as continuous (e.g. age). The thresholds would be computed by setting this variable to the benchmark level for all individuals in the sample.

are only due to genuine heterogeneity in the outcome of interest (in our case, work-disability) and no longer to heterogeneity in reporting styles.

These exercises are the so-called “counterfactuals” and they may answer key questions like: how many people in Country X would report to be work disabled if respondents of Country X used the response scales of respondents of Country Y?

In this case, if the threshold equations include a dummy variable for Country Y, we estimate the benchmark scale by predicting the thresholds under the assumption that all respondents live in Country Y and then the corresponding dummy is always equal to 1.

4 Results

In the empirical analysis we control for a large set of explanatory variables, which include demographics (gender and a quadratic polynomial in age), education, employment, marital status, mental and physical health status (grip strength, body mass index, presence of arthritis, mobility, limitations with activities of daily living and instrumental activities of daily living, etc.) and cognitive abilities, as well as country dummies (see Table 1 for more detailed information on each of these variables). Note that for health we only include objective indicators, since subjective questions on health might also be affected by the reporting bias.

The results of our estimation are reported in Table 2. The first column shows the estimates obtained using a baseline model that does not allow for any threshold variation across respondents, the second column reports the estimates for the self-assessment component of the Hopit model, while the results for the threshold equations of the Hopit model are presented in columns 3–6.

The baseline specification in column 1 is nested in the more general Hopit model since it sets to 0 all the parameters but the constant in the threshold equations. This restricted model is almost identical to a standard ordered probit model in that it does not take into account potential differences in reporting styles. A formal likelihood-ratio test strongly rejects the restricted model not allowing for response scale variation against the more general model that does allow for correction of the DIF bias.³ Therefore, in what follows we focus only on the results obtained with the Hopit model.

The estimates in column 2 show that the major determinants of work disability are health conditions. Apart from being currently a smoker, all physical and menve work limiting health conditions than women, while the other socio-demographic variables do not seem to play a major role in explaining the variability in perceived work-disability across individuals. It is interesting to note that, even after controlling for socio-economic characteristics and health indicators, there is still significant cross-country heterogeneity in work-disability rates. In particular, respondents living in Mediterranean countries are those associated with the lowest perceived levels of work-disability.

The estimates in columns 3–6 of Table 2 show that the thresholds significantly depend on a number of variables, such as country dummies, age, education, employment and several health conditions. In order to show the importance of controlling for heterogeneity in reporting styles, we carry out a counterfactual exercise as described in Sect. 3. Figure 3 reports the percentage of individuals who would rate themselves as work disabled if they used, respectively, the reporting styles of the Spanish and of the Dutch respondents.

³ The test statistics is asymptotically distributed as a chi-squared distribution with 108 degrees of freedom and in our sample it is equal to 1,637.588, p value = 0.000.

Table 2 Hopit model, determinants of work disability—2004 wave, country dummies, 9 vignettes. The first column refers to a baseline Hopit specification not allowing for threshold variation across individuals

	Baseline	Self assessments	Threshold equations			
			γ_1	γ_2	γ_3	γ_4
SE	0.005 (0.112)	-0.161 (0.118)	-0.107* (0.055)	-0.043 (0.072)	-0.245*** (0.046)	-0.077* (0.044)
NL	-0.265** (0.104)	-0.298*** (0.110)	-0.145*** (0.049)	0.428*** (0.053)	-0.123*** (0.039)	-0.369*** (0.047)
BE	-0.152 (0.101)	-0.166 (0.106)	-0.111** (0.047)	0.294*** (0.054)	-0.047 (0.038)	-0.185*** (0.045)
FR	-0.449*** (0.097)	-0.422*** (0.102)	0.031 (0.043)	-0.028 (0.055)	0.006 (0.035)	0.056 (0.040)
ES	-0.368*** (0.116)	-0.770*** (0.124)	-0.630*** (0.067)	0.557*** (0.066)	-0.043 (0.044)	0.116** (0.046)
IT	-0.556*** (0.115)	-0.499*** (0.121)	0.001 (0.049)	0.254*** (0.058)	-0.089** (0.042)	-0.399*** (0.052)
GR	-1.290*** (0.114)	-1.389*** (0.120)	-0.116** (0.046)	0.017 (0.058)	-0.146*** (0.038)	-0.233*** (0.041)
Male	0.562*** (0.079)	0.517*** (0.082)	-0.033 (0.035)	-0.037 (0.040)	0.026 (0.029)	0.007 (0.032)
Age	-0.104* (0.059)	-0.055 (0.062)	0.061** (0.026)	-0.048* (0.029)	0.045** (0.022)	0.045* (0.024)
Age2	0.000 (0.089)	0.018 (0.093)	0.019 (0.035)	0.014 (0.037)	-0.011 (0.032)	0.021 (0.032)
Med_educ	-0.023 (0.066)	-0.108 (0.069)	-0.134*** (0.029)	0.145*** (0.032)	0.017 (0.024)	0.054** (0.027)
High_educ	-0.055 (0.072)	-0.121 (0.076)	-0.108*** (0.031)	0.132*** (0.033)	0.009 (0.026)	0.040 (0.028)
Ever_worked	-0.136 (0.101)	-0.112 (0.106)	-0.023 (0.042)	0.118** (0.052)	0.003 (0.038)	-0.029 (0.039)
Partner	-0.028 (0.069)	-0.014 (0.072)	0.015 (0.029)	0.021 (0.034)	0.012 (0.025)	0.037 (0.027)
Overweight	0.112* (0.059)	0.109* (0.062)	0.011 (0.026)	-0.038 (0.029)	0.020 (0.021)	-0.027 (0.023)
Obese	0.258*** (0.074)	0.318*** (0.077)	0.140*** (0.032)	-0.132*** (0.038)	-0.062** (0.029)	-0.130*** (0.031)
Smoking	-0.055 (0.060)	-0.027 (0.063)	0.043* (0.026)	-0.032 (0.029)	0.052** (0.022)	-0.025 (0.024)
Arthritis	0.361*** (0.072)	0.356*** (0.075)	-0.007 (0.034)	0.016 (0.038)	-0.008 (0.029)	-0.004 (0.033)
Chronic	0.375*** (0.060)	0.400*** (0.063)	-0.005 (0.028)	0.043 (0.030)	-0.000 (0.023)	0.011 (0.026)
Symptoms	0.376*** (0.061)	0.336*** (0.064)	-0.107*** (0.030)	0.152*** (0.033)	0.072*** (0.024)	-0.039 (0.027)

Table 2 continued

	Baseline	Self assessments	Threshold equations			
			γ_1	γ_2	γ_3	γ_4
Mobility	0.482*** (0.059)	0.464*** (0.062)	-0.014 (0.027)	-0.025 (0.030)	-0.009 (0.023)	0.069*** (0.025)
ADL	0.462*** (0.114)	0.430*** (0.119)	-0.123** (0.061)	0.158*** (0.061)	-0.004 (0.047)	-0.119** (0.055)
IADL	0.323*** (0.093)	0.369*** (0.097)	0.072 (0.045)	-0.067 (0.050)	0.025 (0.038)	0.001 (0.044)
Eurocat	0.324*** (0.064)	0.316*** (0.067)	-0.005 (0.030)	-0.005 (0.033)	-0.038 (0.025)	0.028 (0.028)
Fluency	-0.126*** (0.045)	-0.110** (0.047)	0.024 (0.019)	-0.025 (0.022)	-0.029* (0.017)	0.021 (0.018)
Recall	-0.012 (0.018)	-0.041** (0.018)	-0.041*** (0.008)	0.039*** (0.009)	-0.001 (0.006)	-0.011 (0.007)
Grip	-0.178*** (0.032)	-0.178*** (0.033)	-0.009 (0.015)	0.032* (0.017)	-0.009 (0.012)	0.005 (0.013)
Constant			-0.143 (0.203)	-0.829*** (0.100)	-0.138* (0.079)	-0.042 (0.085)
Vignette parameters						
θ_1	1.481*** (0.201)	1.254*** (0.209)				
θ_2	0.763*** (0.199)	0.546*** (0.208)				
θ_3	1.488*** (0.201)	1.267*** (0.209)				
θ_4	0.877*** (0.200)	0.662*** (0.208)				
θ_5	0.956*** (0.200)	0.736*** (0.208)				
θ_6	0.534*** (0.199)	0.312 (0.208)				
θ_7	1.190*** (0.200)	0.971*** (0.208)				
θ_8	1.191*** (0.200)	0.972*** (0.208)				
θ_9	1.653*** (0.202)	1.435*** (0.210)				
σ_v^2	0.505	0.464				
Log-likelihood	-30,515.938	-29,697.144				
Observations	2,464	2,464				

Standard errors in parentheses, *** p value < 0.01, ** p value < 0.05, * p value < 0.1

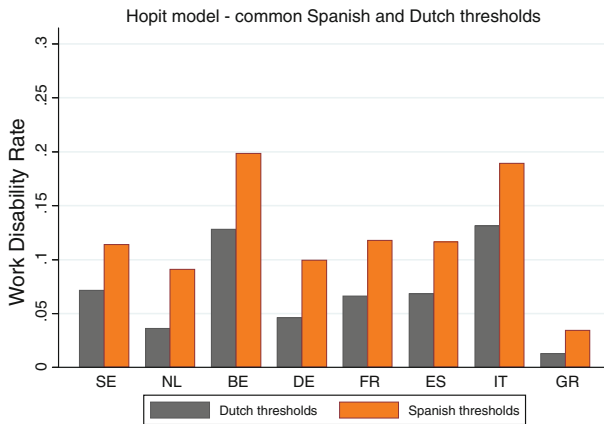


Fig. 3 Counterfactual—Spanish and Dutch thresholds

We focus on Spain and The Netherlands because Fig. 2 suggests that their respondents adopt different response scales when answering vignette questions. Indeed, Spanish respondents are more likely to consider the hypothetical person described in the vignettes as work-disabled. In addition, if the institutional background affects the concept of work-disability underlying self-reports, it is worth comparing these two countries because they are associated with different architectures of work-disability insurance, as suggested by the evidence proposed in Börsch-Supan (2005). Figure 3 confirms once again that response scales do play a role. When the Dutch thresholds are imposed, the resulting work-disability rates in all countries are systematically lower than those obtained using the Spanish thresholds. As an example, the work-disability rate in France is about 6% according to the Dutch reporting styles but it increases to 11% when the Spanish reporting style is adopted. Given the same perceived level of work disability Y^* , Spanish respondents are more likely to declare the presence of work limiting health problems than their Dutch counterparts.

4.1 The Role of Institutions

In line with the results of Table 2, Fig. 3 reveals that, even when we correct for the DIF bias, we still find large cross-country variation in work-disability rates. To understand to what extent this heterogeneity is due to institutional differences, we re-estimate the same specifications as before replacing the country dummies with three macro-economic indicators released at the country level by Eurostat for the 2004 year: the Harmonised Index of Consumer Prices, the employment rate for the population in the age group 55–64 and the public expenditure for disability function as a percentage of the national GDP. The first index is an indicator of inflation and price stability of the European Central Bank: it reflects cross-country variability in the price of goods and services, including those associated with the presence of disability, such as health care. The employment rate for the individuals aged 55–64 intends to summarize the characteristics of the labour market policies targeted to the elderly as well as the incentives to remain at work provided by Social Security in terms of eligibility requirements for pension benefits and replacement rates. Finally, the public expenditure for the disability function measures the generosity of the welfare state in maintaining the income of work-disabled persons. Our results are reported in Table 3. Overall, the parameters on socioeconomic characteristics and objective health indicators

are unaffected by replacing the country dummies with the macro-economic indicators. The DIF bias matters even in this case and a formal likelihood-ratio-test still shows that the hypothesis of common thresholds across individuals is not supported by the data.⁴ Focusing on the relationship between work-disability and macro-economic indicators, we find that the employment rates for individuals aged 55-64 and the generosity of the disability function are two relevant predictors of the perceived level of work-disability. Indeed, respondents in countries with higher employment rates are less likely to be work-disabled. This result might suggest that introducing active policies stimulating the labour market attachment of the elderly results in a lower propensity towards opting out of the labour force via the enrollment in disability programs, which is often a pathway to retirement. On the contrary, a more generous disability function is associated to higher work disability rates. Our results support the hypothesis that the cross-country differentials found in Table 2 and Fig. 3 are related to macro-economic indicators and that countries providing older individuals with higher chances of employment and less generosity in the disability schemes are those where individuals are less likely to report work limiting health problems.

Further, the estimates in Table 3 show that employment rates and the generosity of disability schemes affect individual reporting styles for work-disability. One possible explanation for this result is that more generous disability insurance schemes can induce individuals to classify themselves as work-disabled even in presence of mild limitations. In Fig. 4 we represent the percentage of respondents in each country that would rate themselves as work disabled if they were using the reporting styles of the countries with the least and the most generous disability benefits, respectively Greece and Sweden. Interestingly, we find that more generous disability benefits would induce respondents of all countries to consider the same health problems as more severe, *ceteris paribus*.

5 Conclusion

The proportion of elderly individuals in European societies is steadily rising and this demographic trend poses the necessity of designing a welfare state able to foster their social and economic inclusion. Therefore, a thorough understanding of the composition of the disabled population helps the development of policies aimed at (1) reducing and preventing the presence of health impairments limiting working capacity, (2) supporting the rehabilitation of disabled persons and, finally, (3) stimulating the participation of this population group to the economic activities of their countries. However, how to measure work disability in practice is a quite complicated task. Physical measures of health can be expensive and often are related to specific features of a given health domain. Moreover, they provide information on health rather than working capacity. Self-reported measures of work disability are, therefore, the common solution adopted by researchers. In fact, several surveys ask respondents to self-assess their work-disability status according to a pre-determined multi-item scale. The relationship between this indicator of work-disability and the explanatory factors of interest is then studied via standard ordered probit techniques. However, subjective judgements might be biased.

⁴ As before, the likelihood-ratio test compares the baseline specification in the first column with the Hopit model. The test statistics is asymptotically distributed as a chi-squared distribution with 92 degrees of freedom and in our sample it is equal to 1,078.03, p value = 0.000.

Table 3 Hopit model, determinants of work disability—2004 wave, macro-indicators, 9 vignettes. The first column refers to a baseline Hopit specification not allowing for threshold variation across individuals

	Baseline	Self assessments	Threshold equations			
			γ_1	γ_2	γ_3	γ_4
Ind_prices	0.054 (0.149)	-0.160 (0.157)	-0.141* (0.073)	-0.230*** (0.086)	0.032 (0.058)	0.431*** (0.065)
Emp_5564	-2.920*** (0.791)	-3.410*** (0.831)	-0.407 (0.355)	-0.203 (0.399)	-0.613** (0.289)	-1.101*** (0.339)
Dis_funct	0.429*** (0.124)	0.355*** (0.130)	-0.039 (0.058)	-0.095 (0.068)	0.026 (0.047)	0.275*** (0.053)
Male	0.515*** (0.078)	0.481*** (0.081)	-0.032 (0.036)	-0.023 (0.043)	0.013 (0.029)	-0.022 (0.031)
Age	-0.081 (0.058)	-0.037 (0.061)	0.042 (0.026)	-0.031 (0.031)	0.053** (0.022)	0.075*** (0.023)
Age2	-0.016 (0.087)	-0.011 (0.092)	-0.002 (0.037)	0.042 (0.041)	-0.017 (0.032)	0.015 (0.032)
Med_educ	0.029 (0.063)	-0.041 (0.066)	-0.099*** (0.029)	0.059* (0.033)	0.042* (0.023)	0.103*** (0.026)
High_educ	-0.060 (0.071)	-0.119 (0.074)	-0.078** (0.032)	0.061* (0.036)	0.021 (0.026)	0.052* (0.028)
Ever_worked	-0.105 (0.099)	-0.081 (0.104)	-0.014 (0.043)	0.100* (0.057)	0.026 (0.037)	0.036 (0.039)
Partner	-0.007 (0.068)	-0.005 (0.072)	-0.012 (0.030)	0.060 (0.037)	0.012 (0.025)	0.038 (0.027)
Overweight	0.099* (0.058)	0.097 (0.062)	0.015 (0.027)	-0.054* (0.031)	0.025 (0.021)	-0.019 (0.023)
Obese	0.236*** (0.073)	0.299*** (0.077)	0.127*** (0.033)	-0.128*** (0.041)	-0.063** (0.029)	-0.119*** (0.031)
Smoking	-0.102* (0.059)	-0.078 (0.062)	0.030 (0.026)	-0.016 (0.031)	0.039* (0.021)	-0.059** (0.024)
Arthritis	0.353*** (0.071)	0.369*** (0.074)	0.028 (0.035)	-0.024 (0.041)	-0.001 (0.029)	-0.005 (0.033)
Chronic	0.373*** (0.060)	0.390*** (0.063)	0.005 (0.028)	0.033 (0.033)	-0.001 (0.023)	0.005 (0.026)
Symptoms	0.379*** (0.061)	0.346*** (0.064)	-0.087*** (0.031)	0.125*** (0.035)	0.078*** (0.024)	-0.018 (0.027)
Mobility	0.476*** (0.059)	0.455*** (0.062)	-0.017 (0.027)	-0.029 (0.032)	-0.011 (0.023)	0.069*** (0.025)
ADL	0.455*** (0.114)	0.414*** (0.118)	-0.107* (0.063)	0.136** (0.069)	0.005 (0.047)	-0.107* (0.055)
IADL	0.339*** (0.092)	0.375*** (0.096)	0.040 (0.046)	-0.019 (0.055)	0.019 (0.038)	0.000 (0.043)
Eurocat	0.306*** (0.063)	0.310*** (0.066)	0.025 (0.031)	-0.049 (0.037)	-0.025 (0.025)	0.039 (0.027)

Table 3 continued

	Baseline	Self assessments	Threshold equations			
			γ_1	γ_2	γ_3	γ_4
Fluency	-0.092** (0.044)	-0.073 (0.046)	0.026 (0.019)	-0.032 (0.023)	-0.016 (0.017)	0.054*** (0.018)
Recall	-0.027 (0.017)	-0.049*** (0.018)	-0.028*** (0.008)	0.020** (0.009)	-0.005 (0.006)	-0.020*** (0.007)
Grip	-0.164*** (0.031)	-0.169*** (0.033)	-0.012 (0.015)	0.030* (0.018)	-0.006 (0.012)	0.011 (0.013)
Constant			-0.168 (0.412)	0.277 (0.235)	-0.142 (0.160)	-1.378*** (0.173)
Vignette parameters						
θ_1	1.781*** (0.406)	0.843** (0.425)				
θ_2	1.080*** (0.404)	0.135 (0.425)				
θ_3	1.788*** (0.406)	0.849** (0.425)				
θ_4	1.191*** (0.405)	0.249 (0.425)				
θ_5	1.269*** (0.405)	0.325 (0.425)				
θ_6	0.856** (0.404)	-0.094 (0.425)				
θ_7	1.496*** (0.405)	0.557 (0.425)				
θ_8	1.498*** (0.405)	0.559 (0.425)				
θ_9	1.949*** (0.406)	1.017** (0.425)				
σ_v^2	0.482	0.469				
Log-likelihood	-30,560.970	-30,021.955				
Observations	2,464	2,464				

Standard errors in parentheses, *** p value < 0.01, ** p value < 0.05, * p value < 0.1

Whenever we aim at comparing self-assessments across countries, we should take into account the fact that individuals in different countries can interpret or understand the same question in different ways, because they use different scales to evaluate themselves. These inter-personal and inter-cultural differences in interpreting, understanding or using response categories for the same question is called *differential item functioning* (DIF).

The vignette methodology developed by King et al. (2004) is a generalization of the common ordered probit model, where DIF is modelled through variations in the thresholds. The thresholds determine the individual response scales and are allowed to vary with individual characteristics and across countries.

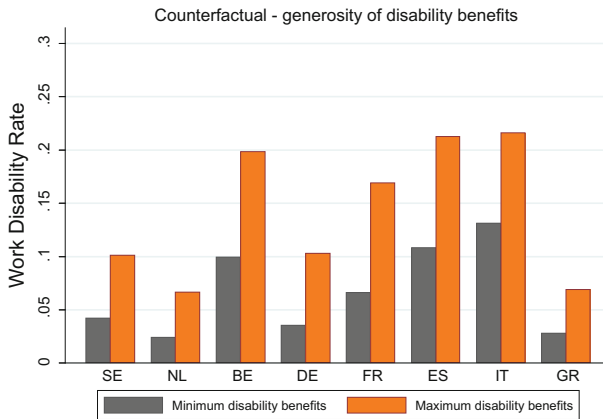


Fig. 4 Counterfactual—the effect of the generosity of disability benefits

Using vignette data from the 2004 wave of the Survey of Health, Ageing and Retirement in Europe (SHARE), the empirical analysis in this paper focuses on individuals aged 50–64 in eight European countries and looks at the determinants of work-disability differentials across countries. Indeed our data show that in seemingly-similar European countries there are wide raw differences in the proportions of individuals who declare to be work-disabled. We first estimate cross-country heterogeneity in work-disability by conditioning on a wide set of socioeconomic and health indicators collected at the individual level. Our findings show that cross-country differentials in work-disability are still sizeable even after conditioning on our set of control factors. Moreover, a formal likelihood-ratio test confirms that DIF is an issue in our sample and the data do not support the hypothesis that reporting styles (i.e. the thresholds) are invariant across individuals. In particular, reporting styles are affected by the country of residence, age, education, employment and health status.

To better understand the role of institutions in explaining the cross-country heterogeneity in work-disability rates we analyze how they are influenced by a set of macroeconomic indicators released by Eurostat for each country in our sample: the Harmonized Index of Consumer Prices, the employment rate for individuals aged 55–64 and public expenditure for disability function as percentage of the national GDP. We find that respondents living in countries with higher employment rates and less generous disability schemes are less likely to suffer work limitations. This evidence seems to suggest that policies stimulating the labour market attachment of the elderly and tightening the requirements of disability benefits might actually reduce the financial burden on disability insurance schemes that question their long-run sustainability.

Finally, we find that institutions affect work-disability reporting by modifying individuals' reporting styles. A counterfactual simulation shows that individuals living in countries with more generous disability schemes are more likely to declare to be work-disabled. This result informs the literature on work-disability by showing that evaluating the impact of policies only on the basis of self-evaluations might lead to misleading results: institutional changes can affect not only the true work-disability status of the elderly but also their response scales used in providing self-evaluations.

Acknowledgments We are grateful to the editors and two anonymous referees for helpful comments and suggestions. Financial support from the European Commission through the STREP project COMPARE n. 028857 is gratefully acknowledged. This paper uses data from SHARE Wave 1, as of November 2009. SHARE data collection in 2004–2007 was primarily funded by the European Commission through its 5th and 6th framework programmes (project numbers QLK6-CT-2001- 00360; RII-CT- 2006-062193; CIT5-CT-2005-028857). Additional funding by the US National Institute on Aging (grant numbers U01 AG09740-13S2; P01 AG005842; P01 AG08291; P30 AG12815; Y1-AG-4553-01; OGH04-064; R21 AG025169) as well as by various national sources is gratefully acknowledged (see <http://www.share-project.org> for a full list of funding institutions).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Bago d'Uva, T., Lindeboom, M., O'Donnell, O., & van Doorslaer, E. (2009). *Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity*. HEDG Working Paper 09/30, Health and Econometrics Data Group, The University of York.
- Bonoli, G., & Shinkawa, T. (Eds.). (2005). *Ageing and pension reform around the world. Evidence from eleven countries*. Cheltenham: Edward Elgar.
- Börsch-Supan, A. (2005). Work disability and health. In A. Börsch-Supan, A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist, & G. Weber (Eds.), *Health, ageing and retirement in Europe. First results from the survey on health, ageing and retirement in Europe*. Mannheim: MEA.
- Börsch-Supan, A. (2008). Changes in health status and work disability. In A. Börsch-Supan, A. Brugiavini, H. Jürges, A. Kapteyn, J. Mackenbach, J. Siegrist, et al. (Eds.), *First results from the survey of health, ageing and retirement in Europe (2004–2007). Starting the longitudinal dimension* (pp. 230–238). Mannheim: MEA.
- Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A., Mackenbach, J., Siegrist, J., et al. (Eds.). (2008). *First results from the survey of health, ageing and retirement in Europe (2004–2007). Starting the longitudinal dimension*. Mannheim: MEA.
- Börsch-Supan, A., Brugiavini, A., Jürges, H., Mackenbach, J., Siegrist, J., & Weber, G. (Eds.). (2005). *Health, ageing and retirement in Europe. First results from the survey on health, ageing and retirement in Europe*. Mannheim: MEA.
- Burkhauser, R. V., & Daly, M. C. (2002). Policy watch: U.S. disability policy in a changing environment. *Journal of Economic Perspectives*, 16(1), 213–24.
- Datta Gupta, N., Kristensen, N., & Pozzoli, D. (2010). External validation of the use of vignettes in cross-country health studies. *Economic Modelling*, 27, 854–865.
- Economic Policy Committee. (2000). *Progress report of the ecofin council on the impact of ageing populations on public pension systems*. Economic Policy Committee, EPC/ECFIN/581/00-ENRev. 1, Brussels.
- Gruber, J., & Wise, D. (Eds.). (2004). *Social security programs and retirement around the world: Micro-estimation*. Chicago, IL: University of Chicago Press.
- Haveman, R., & Wolfe, B. (2000). The economics of disability and disability policy. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (Vol. 1B, pp. 995–1051). Amsterdam: Elsevier.
- Kapteyn, A., Smith, J., & van Soest, A. (2007). Vignettes and self-reports of work disability in the United States and the Netherlands. *American Economic Review*, 97(1), 461–473.
- King, G., Murray, C., Salomon, J., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 191–207.
- King, G., & Wand, J. (2007). Comparing incomparable survey responses: New tools for anchoring vignettes. *Political Analysis* 15(1), 46–66.
- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling Using Stata* (2nd ed.). Texas: STATA Press.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). *GLLAMM manual*. U.C. Berkeley Division of Biostatistics Working Paper Series, Working Paper 160.
- Van Soest, A., Kapteyn, A., Andreyeva, T., & Smith, J. (2006). *Self-reported disability and reference groups*. W-409, RAND labour & population, working paper series.
- Van Soest, A., Delaney, L., Colm, H., Kapteyn, A., & Smith, J. (2007). *Validating the use of vignettes for subjective threshold scales*. IZA discussion paper 2860, Institute for the Study of Labor, Bonn.