UCD GEARY INSTITUTE
DISCUSSION PAPER
SERIES

Work Disability, Work Justification Bias in Europe and the US

Work Disability, Work, and Justification Bias in Europe and the U.S.

Arie Kapteyn, James P. Smith, RAND
Arthur van Soest, Tilburg University, Netspar & RAND

August 2009

Abstract

To analyze the effect of health on work, many studies use a simple self-assessed health measure based upon a question such as "do you have an impairment or health problem limiting the kind or amount of work you can do?" A possible drawback of such a measure is the possibility that different groups of respondents may use different response scales. This is commonly referred to as "differential item functioning" (DIF). A specific form of DIF is justification bias: to justify the fact that they don't work, non-working respondents may classify a given health problem as a more serious work limitation than working respondents. In this paper we use anchoring vignettes to identify justification bias and other forms of DIF across countries and socio-economic groups among older workers in the U.S. and Europe. Generally, we find differences in response scales across countries, partly related to social insurance generosity and employment protection. Furthermore, we find significant evidence of justification bias in the U.S. but not in Europe, suggesting differences in social norms concerning work.

1. Introduction

The fraction of workers on disability insurance (DI) is vastly different across Western European countries with similar levels of economic development and comparable access to modern medical technology and treatment (Eurostat, 2001). Institutional differences in eligibility rules or generosity of benefits contribute to an explanation of differences in disability rolls (Boersch-Supan, 2007). Recent survey data show, however, that significant differences between Western European countries and the U.S. are also found in self-reports of work limiting disabilities. Table 1 below, taken from Kapteyn et al. (2009), illustrates the point.[1] We see considerable variation in both DI-expenditures as a percentage of GDP and the percentage of males between 40 and 65 reporting some form of work disability. Remarkably the two columns are only weakly correlated: the correlation is 0.20.

Table 1: Expenditures on Disability Insurance and

Self-reported Male Work Disability, 2001

	DI expenditure as a % of GDP	Self-reported male work disability, 40-65, 2001 (%)
Germany	1.6	40.3
Denmark	2.7	22.0
Netherlands	4	24.5
Belgium	2.2	14.3
France	1.7	20.5
UK	2.2	13.1
Ireland	1.3	15.7
Italy	2	8.0
Greece	1.6	13.3
Spain	2.3	15.5
Portugal	2.4	22.9
Austria	2.3	17.8
Finland	3.1	29.0
U.S.	1.1	19.3

---

[1] Self-reports for the European countries are taken from the European Community Household Panel. For the U.S. they are taken from the Panel Study of Income Dynamics. The exact question on work disability in ECHP is: "Are you hampered in your daily activities by any physical or mental health problem, illness or disability?" In the PSID, it is: "Do you have any physical or nervous condition that limits the type of work or the amount of work you can do?"

Croda and Skinner (2009), using data from SHARE and HRS, find little or no evidence that in countries with a higher proportion of individuals on DI, the fraction of DI recipients with self-reported fair or poor health is lower. This is what one would expect if access to DI were largely driven by health considerations and if the distribution of health in different countries would be roughly the same. In that case a system with strict eligibility rules or less generous benefits would mainly select individuals with the worst health condition. The more generous benefits or eligibility rules become, the more likely it is that individuals in better health are drawn into the pool of DI recipients. One explanation for the weak relation between DI-recipiency rates and self-reported health across countries may lie in differences in response scales used to answer subjective health questions adopted by residents of different countries.

This paper applies the same vignette methodology as in Kapteyn, Smith, and Van Soest (2007) to determine the extent to which differences in self-reported work-limiting disability between several European countries and the U.S. are due to differences in response scales. In addition, we also consider the effect of work disability on employment and the potential effect of justification bias on the estimate of this effect.

The remainder of the paper is organized as follows. The second section describes the data collection efforts in HRS and SHARE that make this research possible. Section 3 outlines the vignette methodology and our statistical model that corrects for response scale differences across countries. The fourth section briefly describes the vignettes used in this study and how respondents in different European countries and the US respond to the same vignette scenarios. Section 5 presents the empirical results and their implications for interpreting observed differences in self-reported work disability. Section 6 extends the model developed in Section 3 to include employment and justification bias. Section 7 presents the estimation results of that

model and shows their economic importance by means of a number of simulations. Our principal conclusions are contained in the last section.

2. Data Sources

For the United States, we use the 2004 wave of the Health and Retirement Study (HRS), a bi-annual panel with a representative sample of the US population over 50. It has been conducted by the University of Michigan since1992. Information collected includes physical and mental health, socio-economic status (including measures of labor market status, income, education and wealth), social support, etc. The surveys use a mixture of modes with most new interviews conducted face-to-face and most re-interviews by phone. Wave-specific overall response rates for the HRS have improved from 81.7% in 1992 to close to 90% at later waves, specifically 87.8% in 2004. The survey has a complex sample design and oversamples Blacks, Hispanics, and residents of Florida. Details on the HRS methodology and the 2004 Wave are available elsewhere (Heeringa and Connor, 1995). For our analysis we use a subsample of respondents who first completed a face-to-face interview and later completed a leave-behind questionnaire. The leave-behind questionnaire consisted of a series of work disability vignettes and was targeted toward respondents less than 75 years of age.

SHARE is a large-scale project that aims to collect interdisciplinary longitudinal data on European citizens of age 50 and older and their spouses. The eleven participating countries in the baseline wave were Denmark, Sweden, Germany, Belgium, the Netherlands, Austria, Switzerland, Spain, Italy, Greece and France.[2] Using a common instrument, SHARE includes information on physical and mental health, socio-economic status including measures of income,

---

[2] The Czech Republic, Poland, Israel, and Ireland, were added later. See www.share-project.org for exact sample sizes.

education and wealth, and social support. The first wave of the main survey was fielded in 2004, with between 1000 and 4000 individuals in each country, adding up to about 29,000 individuals.

While containing many unique features, in order to facilitate additional comparisons with the USA and England, SHARE was purposely modeled after the HRS and the English Longitudinal Survey of Aging (ELSA) and follows a common set-up across all countries with the goal of facilitating cross-country research.

For a subset of countries that agreed to participate, SHARE included a set of self-assessments and vignette questions on general health status and on work limiting disabilities as part of a drop-off questionnaire. The eight countries that participated in this vignette experiment were Germany, France, Spain, Belgium, Greece, Italy, the Netherlands, and Sweden. The work disability vignettes were identical to the work disability vignettes in the HRS leave-behind questionnaire. Both were taken from the surveys used by Kapteyn et al. (2007).

The work disability vignettes deal with work-limiting health problems in the domains of pain, depression, and cardio-vascular disease. An example of a vignette is the following one (the first vignette cited in Appendix B which contains the full set of vignettes):

[Eva] feels worried all the time. She gets depressed once a week at work for a couple of days in a row, thinking about what could go wrong and that her boss will disapprove of her condition. But she is able to come out of this mood if she concentrates on something else.

For each vignette, the respondent is asked: "Please give us your judgment on how limited these people are in the kind or amount of work they can do"[3]

1. Not at all limited
2. Mildly limited
3. Moderately limited
4. Severely limited
5. Cannot do any work

---

[3] This is the question wording in the HRS. The wording in SHARE is slightly different (and translated in each country's language). See www.share-project.org.

The names used in the vignettes (Eva in the example) vary by vignette and, moreover, to each vignette either a male or a female name was randomly assigned. (See Appendix B for details[4])

Preceding the vignette questions, a respondent is asked the self-assessment of work disability question: "To what extent are you limited in the kind or amount of work you can do because of an impairment or health problem?"[5] with the same answer categories as above. As explained in the next section, these vignettes make it possible to analyze cross-country differences in work related health, corrected for international differences in response scales. Moreover, the vignettes will also be used for comparisons of different socio-economic groups.

## 3. The Theory of Vignettes

In this section, we first provide an intuitive description of the use of vignettes for identifying response scale differences and then sketch our statistical approach. The basic idea is illustrated in Figure 1, which presents the distribution of health related work limitations (work disability for short) in two hypothetical countries. The density of the continuous work disability variable in country A is to the left of that in country B, implying that on average, people in country A suffer less work disability than in country B. The people in the two countries, however, use very different response scales if asked to report their work limitations on a five-point scale (none-mild-moderate-severe-extreme). These differences may be caused by cultural differences, or simply be the result of inadequate translation, for instance because there exist no exact one-to-one translations of concepts from one language to the other. In the example in the figure, people in country A have a much more negative view on their capacity for work than people in country B. Someone in country A with the health indicated by the dashed line would report to have a

---

[4] There we also describe some differences in the formulation of the self-reported disability questions and one vignette across the two surveys. In this version of the paper we have ignored these differences.
[5] Again, this is the wording in the HRS; the wording in SHARE is somewhat different: "Do you have any impairment or health problem that limits the kind or amount of work you can do?" with answer categories "None", "Mild", "Moderate", "Severe" and "Extreme."

severe work disability, while a person in country B with the same actual work limitation would report only a mild work disability The frequency distribution of the self-reports in the two countries would suggest that people in country A are more work disabled than those in country B—the opposite of the true disability distribution. Correcting for the differences in the response scales (DIF, "differential item functioning," in the terminology of King et al., 2004) is essential to compare the distributions of actual work limitations in the two countries.

Figure 1. Comparing self-reported health related work limitations
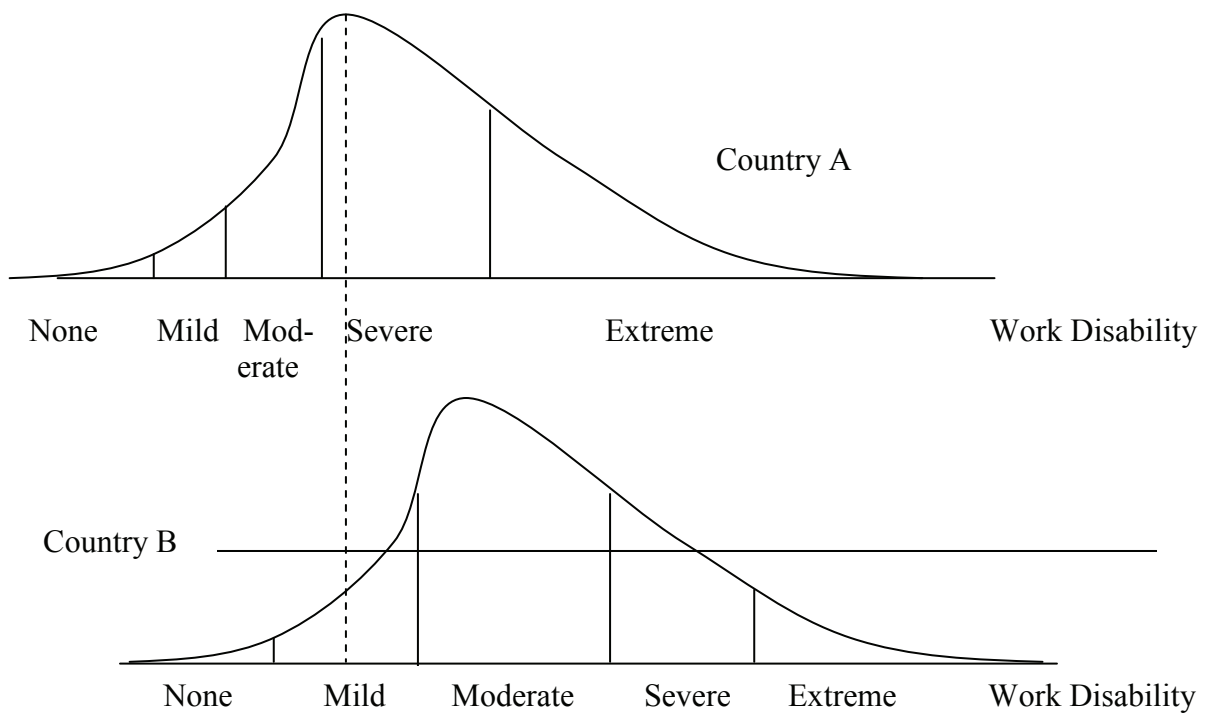
Vignettes can be used to do the correction. A vignette question describes the work limitations of a hypothetical person and then asks the respondent to evaluate the work disability of that person on the same five-point scale that was used for the self-report of their own health. Since the vignette descriptions are the same in the two countries, the vignette persons in the two

countries have the same actual work limitations. For example, respondents can be asked to evaluate the work limitation of a person whose disability is given by the dashed line. In country A, this will be evaluated as "severe." In country B, the evaluation would be "mild." Since the actual work disability is the same in the two countries, the difference in the country evaluations must be due to DIF.

Vignette evaluations thus help to identify differences between the response scales. Using the scales in one of the two countries as the benchmark, the distribution of evaluations in the other country can be adjusted by evaluating them on the benchmark scale. The corrected distribution of the evaluations can then be compared to that in the benchmark country—they are now on the same scale. In the example in the figure, this will lead to the correct conclusion that people in country B are more disabled than those in country A, on average. The main identifying assumption is *response consistency:* a given respondent uses the same scale for self-reports and vignette evaluations. King et al. (2004) provide evidence supporting this assumption by comparing self-reports and vignette evaluations of vision with an objective measure of vision. Van Soest et al. (2007) provide similar supporting evidence by comparing self-reported drinking problems to actual alcohol consumption.

We will apply the vignette approach to work limiting disability, using vignettes not only to obtain international comparisons corrected for DIF, but also for comparisons of different groups within a given country. For example, it is often hypothesized that men self report themselves in better health than objective circumstances would warrant, that as they age people adjust their norms downward about what constitutes good health, and that some of the SES health gradient reflects different health thresholds by SES rather than true health differences. Vignettes offer the potential for systematic testing of these hypotheses.

3.1 A Formal Model of Response Scales and Vignette Corrections

Our model is an extension of the conditional hopit model (Chopit, cf. King et al., 2004, and Kapteyn et al., 2007). It explains respondents' self-reports on work limitations and their reports on work limitations of hypothetical vignette persons. The first of these is the answer Y_i (where i indicates respondent i) to the question

"Do you have any impairment or health problem that limits the kind or amount of paid work you can do?"

The questions on work limitations of the vignette persons use the same 5-point scale and are formulated in the same way ("Does Mr/Mrs X have any impairment or health problem that limits the type or amount of paid work that he/she can do?"). The answers will be denoted by Y_{li} where each respondent i evaluates L vignettes $l=1,...,L$.

Self-reports are modeled as a function of respondent characteristics X_i and an error term ε_i by the following ordered response equation:

$$Y_i^* = X_i\beta + \varepsilon_i; \ \varepsilon_i \ N(0,1), \ \varepsilon_i \text{ independent of } X_i \tag{1}$$

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

The thresholds τ_j^i between the categories are given by

$$\tau_i^0 = -\infty, \ \tau_i^5 = \infty, \ \tau_1^1 = X_i\gamma^1 + u_i, \ \tau_i^j = \tau_i^{j-1} + \exp(X_i\gamma^j), j = 2,3,4 \tag{3}$$

$$u_i \ N(0,\sigma_u^2), \text{ independent of } \varepsilon_i \text{ and } X_i \tag{4}$$

The error term u_i reflects unobserved heterogeneity in the thresholds. The fact that different respondents can use different response scales τ_i^j is called "differential item functioning" (DIF).

Using the self-reports on own work disabilities only, the parameters β and γ^1 cannot be separately identified;[6] the reported outcome only depends on these parameters through their difference. For example, consider country dummies: two people (with the same characteristics) in two different countries can have systematically different work disability, but if the scales on which they report their work disability can also differ across countries, then self-reports alone are not enough to identify the work disability difference between the countries. This was illustrated in Figure 1 above.

For our analysis we are using a common set of $L=9$ vignette questions, three in each of three domains: affect, pain, and heart problems.[7] The evaluations of vignettes $l=1,\ldots, L$ are modeled using similar ordered response equations:

$$Y_{li}^* = \theta_l + \varepsilon_{li} \tag{5}$$

$$Y_{li} = j \text{ if } \tau_i^{j-1} < Y_{li}^* \leq \tau_i^j, \, j = 1,\ldots 5 \tag{6}$$

$$\varepsilon_{li} \sim N(0,\sigma_v^2), \text{ independent of each other, of } \varepsilon_i \text{ and of } X_i \tag{7}$$

The assumption of *"response consistency"* means that the thresholds τ_i^j are the same for the self-reports and the vignettes. The assumption of "vignette equivalence" implies that genuine work-related health of the vignette person Y_{li}^* does not depend on X_i; it only depends on the vignette description (l) and an idiosyncratic error term.[8]

Given these assumptions, it is clear how the vignette evaluations can be used to separately identify β and γ ($=\gamma^1,\ldots\gamma^4$): From the vignette evaluations alone, γ, $\theta_1,\ldots\theta_L$ can be identified (up to the usual normalization of scale and location). From the self-reports, β can then

---

[6] The γ^j for $j>1$ will still be identified.

[7] Kapteyn et al. (2007) also discuss a model in which thresholds are allowed to vary across the three domains of work disability. They find that the results for this more general model are very similar to the results for the model imposing equal thresholds across domains.

[8] Allowing the vignette evaluations to depend upon gender of the vignette person (as was done in Kapteyn et al., 2007) does not affect the results.

be identified in addition. Thus the vignettes can be used to solve the identification problem due to DIF. The two-step procedure is sketched only to make intuitively clear why the model is identified. In practice, all parameters will be estimated simultaneously by maximum likelihood.[9]

Adjusting for DIF is straightforward in this model once the parameters are estimated. Define a benchmark respondent with characteristics $X_i = X(B)$. (For example, choose one of the countries as the benchmark country.) The DIF adjustment would now involve comparing Y_i^* to the thresholds τ_B^j rather than τ_i^j, where τ_B^j is obtained in the same way as τ_i^j but using $X(B)$ instead of X_i. Thus a respondent's work ability is computed using the benchmark scale instead of the respondent's own scale. This does not lead to an adjusted score for each individual respondent (since Y_i^* is not observed) but it can be used to simulate adjusted *distributions* of Y_i for the whole population or conditional on some of the characteristics in X_i. Of course the adjusted distribution will depend upon the chosen benchmark.

4. Responses to Vignettes in Western Europe and the U.S.

Respondents in each of the eight European countries and the U.S. were given vignettes in three domains of work disability—pain, affect, and heart disease. In each of the three domains, three distinct vignettes are used to describe the conditons of a hypothetical person. The actual vignettes used are presented in Appendix B.[10] Table A.1 in Appendix A compares the responses for the pain domain, while Tables A.2 and A.3 do the same for the affect and heart disease (CVD) domains, respectively. The numbering of vignettes corresponds to how vignettes are

---

[9] This is more efficient than the two-step procedure. Since all error terms are independent, the likelihood contribution is a product of univariate normal probabilities over all vignette evaluations and the self-report, which is relatively easy to compute.

[10] The selection of vignettes is the result of simulation studies with the Dutch CentERpanel, where we administered five vignettes per domain and then estimated Chopit models that used subsets of these five. We found that the extreme vignettes (either describing someone who is healthy or someone who is clearly too sick to work) carried little information. We also looked at the effect of the number of vignettes per domain. Obviously, more vignettes will lead to more accuracy, but for practical reasons the number of vignettes per domain had to be limited. The simulation study led to the conclusion that three vignettes per domain strikes a reasonable balance between practical feasibility and statistical accuracy.

presented in Appendix B. For instance, pain1 is the first pain vignette shown in Appendix B, pain2 is the second pain vignette, etc. Although the health conditions of the persons described in the vignettes are supposed to be the same in all countries, Tables A.1-A.3 show that there are large differences in the evaluation frequencies between the countries.

Some differences are striking. For instance pain2 ("[Catherine] suffers from back pain that causes stiffness in her back especially at work but is relieved with low doses of medication. She does not have any pains other than this generalized discomfort.") is said to provide no work limitation by almost 25% of the American respondents, while in the European countries that percentage is less than 4, with very little variation across European countries. Similarly for CVD2 ("[Tom] has been diagnosed with high blood pressure. His blood pressure goes up quickly if he feels under stress. Tom does not exercise much and is overweight.") almost 29% of Americans consider this to be no or only a mild work disability, while the SHARE average is about 16%. The differences for the Affect vignettes are most pronounced. For all three vignettes Americans are much more likely than Europeans to say that the health condition described in the vignette does not constitute a work disability.

Table 2 summarizes the responses to the work disability vignettes in the nine countries. The three vignettes per domain were selected to deliberately eliminate the extremes where individuals in all countries would tend to describe the vignette person as clearly work disabled or clearly not work disabled. As a result, the differences in the scenarios described in the vignettes are often not large (at least among the European countries) and there is noise in the ranking both within and across countries. Table 2 summarizes the vignette evaluations in two ways. First, the five-point scale on work disability is collapsed into three groups, with none and mild combined into one group and severe an extreme into another group. The percentage of respondents

reporting in two of these three groups (none/mild and severe/extreme) is listed for each of the three vignettes in each of the three domains—pain, affect, and CVD.

The second way of arriving at a simple summary is contained in the rows labeled "Average rank" under each domain. This row is derived by ranking the countries in each of the three vignettes within a domain by which country is 'hardest' on the vignette persons—that is which country has the highest percentage of responses in the mild or no work disability category and the lowest percentage of responses in the severe or extreme category. The final row in Table 2 labeled "Grand Average Rank" is just the average of these ranks across the three work disability domains.

Before we concentrate on the patterns in Table 2, we first discuss some correlations summarized in Table 3. The upper panel shows correlations between the average rankings of countries within domains. For example, the entry 0.44 in the upper left corner is the correlation of the average ranking of countries obtained by calculating the average ranking for vignette 1 and for vignette 2 in the pain domain. One would expect that if countries differ strongly in how "soft" or "tough" they are in their vignette evaluations, the ranking of countries would be pretty much the same for each vignette. Table 3 indeed shows that all correlations but one are positive, but often the correlations are not particularly high. The correlations appear lowest for pain and highest for CVD. Thus, in particular for pain there may be a fair amount of noise in the vignette evaluations.

The bottom panel shows the correlation of rank averages across domains. For example, the entry 0.69 is the correlation between the average ranking of countries for pain and for affect, as shown in Table 3. We observe that the correlation of rank averages across domains is much higher than within domains (the average of the three correlations in the bottom panel is 0.67,

while the average of the nine correlations in the top panel equals 0.40). This supports the notion that the sometimes weak correlations in the top panel of Table 3 are at least partly the result of noise. Aggregating across vignettes leads to a considerable reduction of noise and thus to higher correlations in rankings of countries based on different vignettes.

Table 2 indicates several salient patterns. First, residents of the eight European countries do not share a common view on what constitutes a work disability. For example, while a third of the Dutch respondents state that the first CVD vignette constitutes no or only a mild work disability, the comparable fraction for Spaniards is one-in-twenty and for Swedes one in fifty. Yet, the variation within Europe is less striking than the difference between Europe and the US. Comparing the last two columns, which present the US and the SHARE averages, we see that with the exception of pain3 ("[Mark] has pain in his back and legs, and the pain is present almost all the time. It gets worse while he is working. Although medication helps, he feels uncomfortable when moving around, holding and lifting things at work.") European respondents are always more likely to call a vignette person work disabled than Americans.

Second, the ranking amongst the European countries depends to some extent on the specific domain chosen. For example, the Italians are quite demanding ("tough") in the affect domain, but the Dutch are the toughest on CVD. Yet the differences in ranking across domains are not dramatic, consistent with the earlier observed high correlations across domains. The Americans are the toughest overall, followed by the Italians, the Belgians, the Dutch and the French. At the other end of the spectrum, the Greeks, Swedes, and Spaniards appear most inclined to call a health condition work limiting. The Germans are in the middle.

Table 2: Responses to Work Disability Vignettes

	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	US	Total EU
PAIN										
pain 1										
none,mild	10.9	5.5	12.1	13.4	6.6	6	10.7	7.7	10.1	10.1
severe,extreme	54.1	72	61.9	56.4	64.8	80	48.6	60.7	55	58.7
pain 2										
none,mild	29	18.2	31.6	38.1	58.9	10.4	38.6	47.7	77.9	32.8
severe,extreme	16	36.9	20.4	13.9	11.2	54	8.7	10.5	3.4	18.7
pain 3										
none,mild	7.5	1.9	7.8	14.4	4.9	11.4	5.8	7.2	4.9	7.9
severe,extreme	70.4	72.7	68.4	51.9	59.3	51.5	68.5	59.9	74.1	64.5
Average rank	4.8	8.3	4.8	3.0	5.0	6.3	3.8	4.3	4.3	
AFFECT										
Affect 1										
none,mild	23.9	23.7	20.6	39.2	29.3	34	26.9	35	53.3	28.7
severe,extreme	27.5	30.3	35.7	17.6	24.2	16.5	23	23.5	14.5	24.6
Affect 2										
none,mild	28.2	21	22.4	35.6	18.7	14.6	26.2	22.4	55.5	26.9
severe,extreme	22.7	30.6	33.2	16.9	30.4	52.8	24.5	25.7	12.2	25.1
Affect 3										
none,mild	54.8	39.3	43.9	57.2	74.8	9.9	52.9	67.6	80.6	52
severe,extreme	4.5	17.4	15.9	10.7	6.5	57.8	5.8	4.4	3.6	10.8
Average rank	4.7	7.7	7.5	3.2	5.3	7.0	4.7	3.8	1.0	
CVD										
CVD 1										
none,mild	15.9	5.7	14.9	23.7	33.6	1.7	16	22.2	43.3	16.7
severe,extreme	37.5	63.8	51.9	30	27.5	88.5	47.7	31.9	19	43.4
CVD 2										
none,mild	14.6	4.4	5.8	26.2	30.5	26.2	13.6	21.3	28.6	16.6
severe,extreme	41.8	65.9	77	33.7	26.6	26.1	46.9	33.1	30.5	44.5
CVD 3										
none,mild	7.2	3.8	5.1	16.1	18.2	2.3	6.9	10.8	10.7	9.1
severe,extreme	71.3	81.8	85.9	62.4	46.7	86.3	75.8	59.6	64.3	71.1
Average rank	5.5	8.0	7.7	3.2	1.5	6.7	6.2	3.7	2.5	
Grand Average Rank	5.0	8.0	6.7	3.1	3.9	6.7	4.9	3.9	2.6	

Table 3: Correlations between Rankings

Correlations between rankings within domains			
	Pain	Affect	CVD
Vignettes 1 and 2	0.44	0.43	0.51
Vignettes 1 and 3	0.06	0.45	0.89
Vignettes 2 and 3	-0.33	0.72	0.47
Correlations of average rankings across domains			
Correlation between rank averages of pain and affect			0.69
Correlation between rank averages of pain and CVD			0.57
Correlation between rank averages of affect and CVD			0.76

It is of interest to relate national norms about work disability to institutional arrangements. Table 4 provides a very simple way of doing so. The first column reproduces the grand average ranks from Table 2. Columns 2 and 3 show two indicators of employment protection published by OECD (2004). Both indicators aggregate in some fashion three main domains that make it difficult for an employer to dismiss an employee. These domains are (see OECD, 2004, p. 65):

i) difficulty of dismissal, that is legislative provisions setting conditions under which a dismissal is "justified" or "fair";

ii) procedural inconveniences that the employer may face when starting the dismissal process;

iii) notice and severance pay provisions.

Version 2 is somewhat broader than version 1.

There is a fairly strong positive correlation between the strength of employment protection and a country's rank in the vignette distribution. People in countries with more employment protection are on average "softer" on work disability (i.e. more inclined to see a

given health condition as work limiting). Naturally, there are various alternative explanations for this positive correlation. One would be "culture". In a country with a tough culture, citizens are tough on work disability and don't find employment protection very important. This is then reflected in laws with little protection. The US would be a case in point.

Table 4: National Norms about Work Disability and Employment Protection

	Rank	OECD Employment protection indicator	
		version 1	version 2
Italy	3,1	3,1	3,1
Belgium	3,9	2,2	2,5
Netherlands	3,9	2,1	2,3
France	4,9	3	2,9
Germany	5	2,2	2,5
Greece	6,7	2,8	2,9
Sweden	6,7	2,2	2,6
Spain	8	3,1	3,1
US	2,6	0,2	0,7
Correlation with rank	1,00	0,52	0,56

5. Model Estimates for Response Scales and Self reported Work Disability

In this section, we present our parameter estimates for predicting work disability in the nine countries. Our models incorporate a number of standard demographic covariates—age dummies (less than 58, 58-64, 65-71, 72 or more), years of education, dummies for being female and for currently married. In addition, a series of health indicators are included (heart problems, lung disease, high blood pressure, diabetes, pain, arthritis, cancer, CESD score) with for the European countries the EuroD. The benchmark country is the US. We include a full set of interactions with an EU dummy. Furthermore we include country dummies for the separate EU countries. Table 5 lists the means of the explanatory variables and self-reported work disability by country.

Table 5A. Sample Means by Country: Demographics

Country	female	marrlt	educyrs	age 58-64	age 65-71	age72+
USA	0.543	0.693	13.026	0.204	0.227	0.151
Germany	0.558	0.632	13.409	0.229	0.234	0.241
Sweden	0.510	0.669	10.674	0.267	0.147	0.267
Netherlands	0.509	0.726	11.974	0.236	0.193	0.214
Spain	0.545	0.632	6.682	0.226	0.181	0.288
Italy	0.549	0.615	7.239	0.266	0.184	0.261
France	0.558	0.702	9.559	0.221	0.191	0.300
Greece	0.534	0.685	9.205	0.237	0.246	0.241
Belgium	0.535	0.737	10.594	0.202	0.216	0.265

Explanation: female: dummy female; marrlt: dummy married or living together (benchmark: single, divorced, separated, or widowed); educyrs: years of education; age58-64, age65-71, age72+: dummies age groups 58-64, 65-71 and 72 and older, respectively (benchmark: 50-57). All sample means are weighted.

Table 5B. Sample Means by Country: Health Conditions

Country	heart	lung	hbp	diabetes	pain	arthritis
USA	0.192	0.083	0.492	0.160	0.373	0.511
Germany	0.113	0.069	0.343	0.102	0.567	0.131
Sweden	0.121	0.095	0.303	0.081	0.552	0.096
Netherlands	0.099	0.100	0.239	0.081	0.355	0.093
Spain	0.101	0.080	0.343	0.119	0.479	0.239
Italy	0.118	0.131	0.366	0.087	0.591	0.364
France	0.144	0.087	0.288	0.101	0.531	0.311
Greece	0.154	0.050	0.326	0.120	0.503	0.195
Belgium	0.141	0.063	0.286	0.074	0.504	0.256

Country	cancer	obese	cesd	sdis
USA	0.107	0.300	1.302	1.967
Germany	0.063	0.175	0.992	1.994
Sweden	0.085	0.155	0.964	1.983
Netherlands	0.052	0.126	0.909	1.764
Spain	0.058	0.212	1.327	2.028
Italy	0.054	0.152	1.446	1.967
France	0.064	0.158	1.353	1.906
Greece	0.027	0.168	1.166	1.624
Belgium	0.067	0.196	1.128	2.026

Explanation: heart, …cancer: dummies for chronic conditions based upon answers to survey questions "has the doctor ever told you that you have …."; heart: heart problems; lung: lung disease; hbp: high blood pressure; pain: 1 if answer is "yes" to "do you often have pain?"; obese: BMI>30 (based upon self-reported weight and height); cesd: Center for Epidemiological Studies depression score in HRS; 0.5*EURO-D depression score in SHARE; a higher value indicates more depression related symptoms; sdis: self-reported work disability on a scale from 1 (none) to 5 (extreme). All means are weighted.

The first part of the table shows large differences in years of education, with low means in the southern European countries. There are also substantial differences in the age composition, with, for example, relatively few 65-71 year olds in Sweden. Most chronic conditions are much more prevalent in the US than in most European countries. This applies in particular to obesity, high blood pressure, diabetes, arthritis and heart problems. Still, self-reported work disability in the US seems well in line with that in the European countries. Only Greek and to a lesser extent Dutch respondents seem to face fewer work related health impairments. The vignette evaluations suggest that, compared to Europeans, US respondents underemphasize work related health impairments, so that their actual work related health limitations may be larger than those suggested by the means in Table 5B. Simulations based upon the estimates of the econometric model will look at this in detail.[11]

Table A.4 – A.6 in the appendix present estimation results for the complete model, allowing for DIF, as well as for a model without DIF. The latter model is a standard ordered probit explaining the self-reported work disability on a five point scale. The differences between the two models illustrate the effects of allowing for differences in response scales. The model not allowing for response scale variation is strongly rejected by the data, as is immediately apparent from a comparison of the log-likelihoods (at the bottom of Table A.6).

Table A.4 present the estimates of the work disability equation. The estimated coefficients for the demographic attributes yield few surprises. Work disability increases with age, decreases with schooling and is lower for married respondents. Having any of the health conditions that are included in the model makes it more likely that one reports a work disability. The interactions with the dummy for Europe show that in particular the effects of cancer and

---

[11] The difference in wording between the work disability self-assessment in the US and Europe might also play a role.

depression on work disability are stronger in the European countries than in the US. No such effects are found in the model that does not correct for DIF. The country dummies in Table A.4 demonstrate that, once we correct for response scale differences and control for health conditions, there are still substantial differences across European countries, with a particularly low probability to be work disabled in Greece. (The US is the reference category but because of the interactions with the EU dummy, the coefficients cannot be used to compare the European countries with the US; this will be done in the simulations below). The effects are similar, but smaller, in the model without DIF. Both the models with and without DIF show that the effect of education or being married on work disability is less in Europe than in the US.

The significant estimates of the effect of covariates on the thresholds (Table A.5; model with DIF only) explain why the effects of demographics and health conditions on self-reported work disability change once we take response scale differences into account. Comparing effects of country dummies on the first threshold across the European countries, suggests that Italians are toughest and Spaniards softest, consistent with the raw data summarized in Table 2. Compared to country dummies, effects of health conditions and demographics on the first threshold are modest. More educated respondents tend to be softer, albeit less so in the EU than in the US. Females are harder in the US, but not in the EU. Older respondents are softer in the US, but not in the EU.

The interpretation of the estimated parameters for the remaining thresholds is not very straightforward since these parameters reflect shifts of the second threshold relative to the first one (cf. Equation (3)). Rather than extensively discussing these parameters we turn to some simulations that are easier to interpret.

5.1 Simulation Results

Figure 2 provides a first impression of the effects of using different scales. The first column presents the distribution of self-reported work disability for the US, using US data and US parameters. This simply predicts sample observations for the US. Similarly the last column predicts the distribution of work disability using EU (SHARE) data and EU parameters. Comparing the first and last column suggests relatively minor differences in work disability prevalence between the EU and the US. For instance the percentage saying that they have no work disability is predicted to be about 51% in the US and 49% in the EU. On the other hand the percentage of Americans with severe or extreme work disabilities is predicted to be about 16% in the US and only 12% in the EU. These seemingly modest differences are the result of two counter-acting effects: Americans use response scales that are "harder" while Europeans appear to suffer from less work disability. This is illustrated by the remaining three columns in Figure 2.

Figure 2: Simulated work limitations according to different scenarios
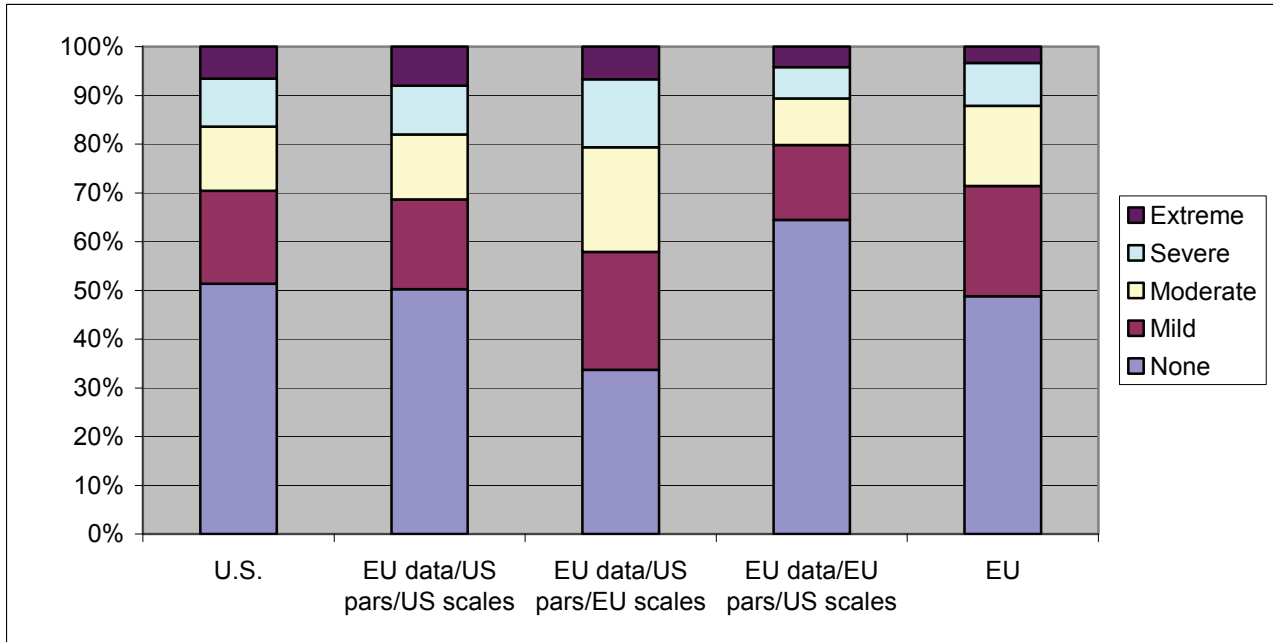
The second column uses EU data, but US scales and US parameter estimates to predict EU disability. The distributions in columns 1 and 2 are virtually identical, with a slight shift in the direction of more work disability in the EU. This implies that distributional differences in demographics and health have only a minor effect on differences in observed work disability prevalence in the US and the European countries.

The middle column is based on EU data, US parameters of the work disability equation and EU response scales. Compared to the first and second column, we now see a dramatic shift in the distribution in the direction of increased work disability. The shift is due to the fact that Europeans use softer response scales and more easily call someone disabled. It is also instructive to make a comparison with the fifth column (both EU data and EU parameters). Since the only difference between the third and fifth column is the use of US parameters in the work disability equation, this comparison suggests that, for given demographics and health conditions, the risk to face an actual work disability is substantially higher in the US than in Europe. For instance,

column 3 implies that 34% has no work limitation, while column 5 implies that 49% has no work limitation.

Finally, a comparison of columns 4 and 5 isolates the effect of response scales on observed work disability distributions in the European countries. When we assign US scales to EU respondents, we observe a dramatic fall in reported work disability. For instance, the percentage of respondents without a work disability goes up from 49% to 64%, while the percentage with severe or extreme work limitations falls from 12% to about 10.5%. Again this is a consequence of the fact that Americans less easily classify someone with a given impairment as work disabled.

Table 6 presents simulated work disability by country. The first row in each panel presents work disability for the US; the second row presents work disability in the country, but using US scales and the third row presents work disability simulated using the response scale of that country.

As was observed for the set of all European countries, using US scales has dramatic effects on the distribution of disability within each European country. It is worth noting that generally most of the changes take place in the categories none and mild. The percentage of respondents in the EU country whose work disability falls in the severe/extreme range is not affected as much. This is as one would expect; there is probably less scope for disagreement about whether a serious health condition constitutes a work disability than whether a mild condition should be seen as work limiting. The biggest reduction in the severe/extreme category takes place in Sweden (from 14.6% to 7.7%) and Spain (from 14.6% to 10.5%). In other countries the changes are much less dramatic and in some cases the adoption of the US scales leads to an increase in the percentage of respondents with severe or extreme work disabilities.

This happens for Italy (an increase by 1.4 percentage points) and The Netherlands (an increase by .3 percentage points).

Table 6: Work Disability by Country

	None	**Mild**	**Moderate**	**Severe**	**Extreme**	**None/ Mild**	**Severe/ Extreme**
Germany							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	61.2	17.7	11.0	6.9	3.2	78.8	10.1
EU	45.4	23.5	18.7	9.8	2.6	68.8	12.5
Sweden							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	68.6	15.1	8.7	5.2	2.5	83.6	7.7
EU	52.2	19.3	14.0	10.5	4.1	71.4	14.6
Netherlands							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	69.6	15.3	8.3	4.8	2.1	84.8	6.8
EU	53.2	29.0	11.3	4.4	2.1	82.2	6.5
Spain							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	68.6	12.8	8.1	5.7	4.8	81.4	10.5
EU	45.6	22.9	16.9	11.1	3.5	68.5	14.6
Italy							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	60.0	15.1	10.3	7.8	6.7	75.1	14.5
EU	47.6	22.9	16.4	7.7	5.4	70.5	13.1
France							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	65.8	15.4	9.2	6.0	3.6	81.2	9.6
EU	52.0	20.6	16.7	8.4	2.2	72.7	10.6
Greece							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	81.5	9.5	4.9	2.7	1.4	91.1	4.1
EU	68.9	15.1	9.4	4.5	2.1	84.0	6.6
Belgium							
US	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, US scales	59.0	17.0	11.3	7.9	4.8	76.0	12.7
EU	43.4	26.7	16.7	8.9	4.4	70.1	13.2

When considering the changes in the categories none and mild we see the biggest

decrease in Spain (by 12.8 percentage points) and Sweden (12.2 percentage points). The smallest

decrease is seen in The Netherlands (2.6 percentage points) and Italy (4.6 percentage points).

We have seen earlier that the use of US scales actually increases the difference in work

disability between the US and the EU countries we are considering here. It is also of interest to

investigate if the use of vignettes reduces the differences between the EU countries. A simple

measure is the variance of the percent none/mild and of the percent severe/extreme across

European countries. We find that the variance of the percentage none/mild is 36.9 if we use the

own country scales and 26.5 if we use the common (US-) scale. For the percentage

severe/extreme, the use of a common scale has virtually no effect. Using the own scale the

variance is equal to 11.0 and when using the common scale the variance equals 10.8.

Table 7: Work Disability by Age in the EU and the US

	50-57			**58-64**			**65-71**			**72plus**			**All**		
	US	EU US, sc.	EU	US	EU US, sc.	EU	US	EU US, sc.	EU	US	EU US, sc.	EU	US	EU US, sc.	EU
None	59.8	75.5	59.4	49.8	71.4	55.5	48.4	61.4	45.9	34.7	48.2	32.9	51.4	64.5	48.8
Mild	17.7	12.7	20.8	19.6	14.2	22.5	19.5	16.2	23.7	21.2	18.6	24.0	19.1	15.3	22.6
Moderate	10.7	6.4	12.5	13.3	7.7	14.0	14.2	10.6	17.4	18.4	14.3	22.5	13.2	9.6	16.4
Severe	7.3	3.7	5.5	10.3	4.6	6.1	10.8	7.3	9.3	14.9	10.5	14.4	9.9	6.4	8.7
Extreme	4.5	1.9	1.8	7.0	2.2	1.9	7.1	4.6	3.7	10.9	8.4	6.3	6.5	4.2	3.4
None/Mild	77.5	88.1	80.3	69.4	85.6	78.1	67.9	77.6	69.6	55.9	66.8	56.9	70.4	79.8	71.4
Sev/Extr.	11.8	5.5	7.2	17.3	6.8	8.0	17.9	11.8	13.0	25.8	18.9	20.7	16.4	10.6	12.1

Next we consider a number of socio-economic categories. Table 7 shows results by age. The scale effects appear to be fairly uniform by age. For instance the effect of going from European scales to US scales on the percentage severe/extreme varies from 1.2 in the 58-64 category to 1.8 in the 72+ category. Similarly the effect on the percentage none/mild varies from -7.5 in the 58-64 category to -9.9 in the 72+ category.

6. Work Disability, Work, and Response Scales

We now extend the model presented in Section 3 by adding an equation explaining employment status, while allowing for justification bias. Justification bias has been introduced by Bound (1991) as a possible effect of non-employment on self reported work disability, where respondents may exaggerate their work limitations to justify that they don't work. The empirical evidence for the existence of justification bias appears to be mixed (see for instance Kreider and Pepper (2007) or Jones (2007) and references therein). The use of vignettes provides a novel approach to the estimation of justification bias.

As before, Y_i denotes the answer of respondent i to the five-point scale self-assessed work disability question while the answers to the work limitations of the vignette persons on the same five point scale are denoted by Y_{li} where each respondent i evaluates L vignettes $l=1,...,L$.

For employment status, we use a binary variable E, with $E_i=1$ if respondent i does some paid work, and $E_i=0$ otherwise. In principle we could distinguish more than two categories here (unemployed, on disability benefits, retired, homemaker, …) but the numbers of observations in some of these employment states are quite small in some countries, making estimation of a richer model difficult.

The equation for self-reported work disability is the same as before, cf. (1)-(2). To keep the model manageable by limiting the number of parameters to be estimated and to facilitate interpretation of the results, we replace the threshold equations (3)-(4) by:

$$\tau_i^0 = -\infty, \ \tau_i^5 = \infty, \ \tau_i^1 = X_i \gamma + \delta E_i + u_i, \ \tau_i^j = \tau_i^{j-1} + \gamma^j, \ j = 2,3,4 \tag{8}$$

$$u_i \ \Box \ N(0, \sigma_u^2), \ \text{independent of } \varepsilon_i \text{ and } X_i \tag{9}$$

This is a more parsimonious specification than the one used in (3). All threshold shifts are the same. The error term u_i again reflects unobserved heterogeneity in thresholds. The effect of employment status E_i on the response scales reflects justification bias: depending on their employment status, respondents may use different response scales. The advantage of the more parsimonious specification is that in this model, justification bias is captured by a single parameter δ (which will be allowed to be different for US and European respondents in the empirical work) instead of a set of parameters affecting several thresholds differently. We expect a positive estimate of δ if employed respondents are less likely to evaluate a given (vignette) person as work disabled than non-employed respondents.

The evaluations of vignettes $l=1,...,L$ are also modeled as before, cf. (5)-(7). Employment status is modeled using a probit-like equation:

$$E_i^* = X_i \varphi + \pi Y_i^* + \eta_i$$
$$\eta_i \ \Box \ N(0,1), \ \eta_i \text{ independent of } X_i, u_i, \varepsilon_{ri}, \varepsilon_{1i},...,\varepsilon_{Li} \tag{10}$$
$$E_i = 1 \text{ if } E_i^* > 0; E_i = 0 \text{ if } E_i^* \leq 0$$

The coefficient π represents the effect of work disability on employment status, which is the main coefficient of interest in many studies of the effect of work related health on employment (cf., e.g., Kreider and Pepper, 2007).

27

In this model, error terms are all independent of each other, and no exclusion restrictions are needed for identification. The intuitive argument for this is as follows: equations (5) and (8) are identified from the vignette evaluations alone, since employment status is exogenous to the errors in (5) and (8). Equation (1) is then identified as well, just like in the model of Section 3. Since the error term ε_i in (1) is independent of the error term in (10), the variation in Y_i^* induced by the variation in ε_i is sufficient to identify equation (10) (although estimation of (10) is complicated by the categorical nature of Y_i).

As a sensitivity check, we will also present results for an over identified version of the model where we impose the exclusion restriction in the employment equation that health conditions affect employment status through work disability only and not directly (maintaining the assumption that all errors are independent of each other).

Figure 3 presents the model including work and justification bias in graphical form. Actual work disability affects employment status (the parameter π in (10)), but we assume there is no reverse effect of employment status on actual work disability, in line with, for example, the results of a recent panel data study of Böckerman and Ilmakkunnas (2009). We could in principle allow for a reverse effect if appropriate exclusion restrictions were available. In our model, employment status affects *reported* work disability through the justification bias effect on response scales (the parameter δ).

The figure does not show the vignette evaluations. We maintain the assumptions underlying the vignette approach discussed in section 3, in particular vignette equivalence and response consistency. The first now also implies that workers and non-workers do not interpret the genuine work related health of the vignette persons in systematically different ways, and we see no reason not to find this assumption plausible. Response consistency now also implies that

justification bias plays the same role in vignette evaluations as in self-assessments. This is an identifying assumption that cannot be tested with the current data. Thus if non-workers evaluate themselves as more work limited than workers with the same actual work related health, they are assumed to also do this with the vignette persons. One might argue that they will be less inclined to exaggerate the vignette person's health problems than their own impairment, which would imply that the effect of employment status on the vignette evaluations is an underestimate of justification bias in self-assessments. Since our estimate of justification bias is driven by the effect on vignette evaluations, this would imply that with our method, we get a lower bound on the "true" justification bias in self-assessments.

Figure 3. Work, work disability, response scales, and justification bias
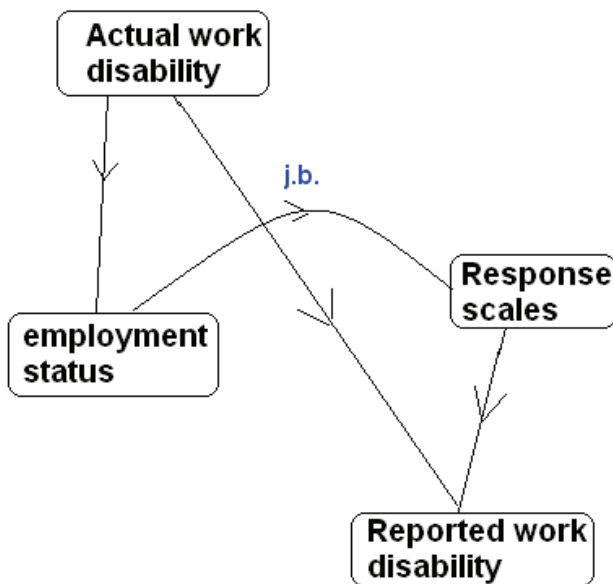
7. Model Estimates for Work Disability, Work, and Threshold Equations

Tables A.7-A.11 present the model estimates with and without DIF. Comparing Table A.7 with the estimates for the disability equation in Table A.4 shows that generally the coefficients are of a similar order of magnitude and have the same sign. One notable difference is the coefficient on gender, which is significantly negative in Table A.7, and which was completely insignificant in Table A.4.

The results for the threshold equation given in Table A.8 cannot be compared directly to Table A.5, as the specification is much simpler. We will use simulations (see below) to see if the simpler specification leads to very different outcomes. The country dummies (with the US as the reference category) appear to be smaller, although Greece maintains its status as somewhat of an outlier. The parameter of primary interest in Table A.8 is the coefficient on the dummy Work. The coefficient is significantly positive in the US, while the interaction with the EU dummy makes clear that the effect of employment status on the thresholds is essentially zero in the EU. In other words, we find significant justification bias in the U.S. and no evidence of justification bias in the EU. It suggests that not working is much more accepted in the EU (for whatever reason) than in the US. Even if our estimate of justification bias is an underestimate of the true justification bias since it does not affect vignette evaluations as much as self-assessments (see the end of the previous section), we still believe this conclusion is justified, since we see no reason why the violation of response consistency would apply in Europe but not in the US. It might mean that justification bias is present in Europe as well as the US, but then it is still much bigger in the US than in Europe.

The estimates of the vignette equation are presented in Table A9. The estimated coefficients on the vignette dummies are similar to those in Table A.6, although they do not

show exactly the same ranking across vignettes. Still, the correlation between the dummy estimates in the two tables is 0.97.

The employment equation estimates are shown in Table A.10. Since the model without DIF is soundly rejected, we concentrate on the estimates for the model with DIF. Since (genuine) work disability is controlled for, the effects of the health dummies indicate the effect of a specific condition on employment *keeping work disability constant*. The negative effects of mental health (measured through the CESD score) and, in the US, diabetes on employment therefore imply that the negative effects of these health conditions on employment are larger than their effect through work disability would suggest. On the other hand, the effect of obesity on employment in the US is smaller than its effect through work disability. The insignificant effects of other health conditions simply mean that the effects of conditions are well captured by the effect of work disability. The estimated age dummies show the expected pattern that employment falls with age. This pattern is much stronger in the EU than in the US. Similarly, females are less likely to be employed than males and also this difference is stronger in the EU than in the US.

The parameter of main interest in the employment equation is the coefficient of work disability. It is more than twice as large in absolute terms in the US than in the EU (-.464 in the US and -.192 in the EU). This result remains essentially unchanged when we specify the employment equations more parsimoniously by omitting the health conditions (Table A.11). One possible interpretation of this is that individuals in the US are more likely to work for pay and that health is one of the main impediments for doing so. In the EU on the other hand, also individuals in relatively good health are often not working, for instance as a result of more generous income replacement schemes. Below we will present simulations that shed more light on this (and other) interpretation(s).

Comparing the coefficient of work disability with the estimate of the same coefficient in a model without DIF shows the effect of controlling for justification bias when estimating the effect of work disability on employment. For the European countries, there is virtually no difference between the DIF corrected and not DIF corrected estimates in Table A.10 (-.192 and -.197, respectively), which corresponds to the finding that there is no evidence of justification bias for these countries. For the US the difference is larger (-.464 versus -.516) and, as expected, we find that the size of the effect of work disability on employment is overestimated if justification bias is not corrected for—part of the negative correlation between work disability and employment is due to the fact that the non-employed tend to over report their work disability. But the size of the bias is modest. (But then again, our estimate might be seen as a lower bound on the bias, if justification bias plays less of a role in vignette evaluations than in self-assessments; see the discussion at the end of Section 6).

7.1 Simulation Results

Figure 4 is analogous to Figure 2 above, but now based on the extended model including an employment equation and allowing for justification bias. Compared to the model underlying Figure 2, the current model has one extra equation (the employment equation) so that there are more possible combinations of data and parameters that can be used in simulating counterfactuals. The first three bars in Figure 4 are most comparable to those in Figure 2, while the last two bars in Figure 4 correspond to the last two bars in Figure 2. Qualitatively the patterns appear to be similar. If we move from US data to EU data (but retaining US scales and parameters), i.e. from the first column to the second, we see again a slight shift in the direction of more work disability in the EU than in the US**.**

When we move from the second to the third column (EU data, US parameters, EU scales) we see an increase in work disability, but the shift is not as dramatic as in Figure 2. Finally, a comparison of the last two columns shows again that if we move from EU data and EU parameters, but US scales to all data, parameters and scales European, then work disability increases substantially (but not as much as in Figure 2). This illustrates once again the difference in response scales used in the EU countries and in the US. The reason that the changes are less dramatic than in Figure 2 is probably the fact that we have imposed that all threshold shifts are parallel to obtain a parsimonious model (see Section 6).

Figure 4: Simulated work limitations; model including employment equation
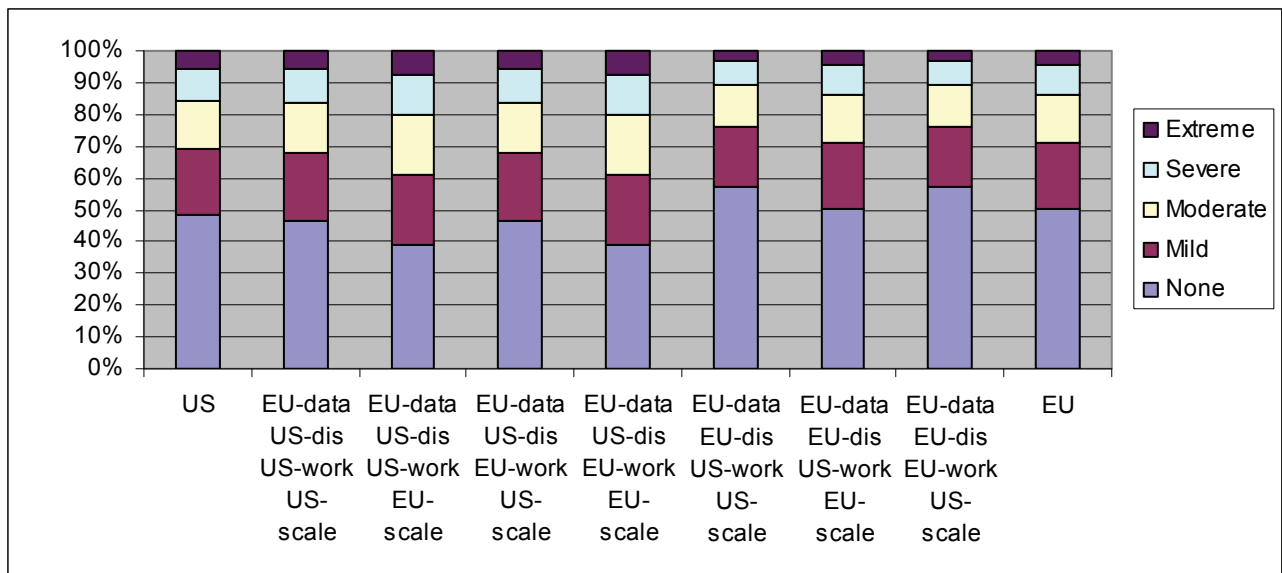
Comparing columns 4 (EU-data, US-disability equation, EU-work equation, US-scale) and 8 (EU-data, EU-disability equation, EU-work equation, US-scale) shows a substantial fall in the prevalence of work disability. Since the only difference between the two columns is the

difference in disability parameters, it illustrates once more the finding that Europeans are less

work disabled than the Americans, although the reported work disability levels are similar.

The only difference between columns 6 (EU-data, EU-disability equation, US-work equation, US-scale) and 8 (EU-data, EU-disability equation, EU-work equation, US-scale) is that in the former the US employment parameters have been used, while in the later we use the EU employment parameters. These differences in parameters appear to have minimal impact on the distribution of observed work disability, even though observed work disability does depend on employment status through its (justification bias) effect on the response scales.

Table 8 and Figure 5 show the percentage of people working for each of the five levels of work disability (the last column of Table 8 is discussed below). As could be expected on the basis of the raw data, a comparison of the first and last row of Table 8 shows that Europeans work less than Americans for any work disability category.

Table 8: Percent Working by Disability Category

	None	Mild	Moderate	Severe	Extreme	Correlations
US	66.6%	45.8%	34.3%	23.5%	12.2%	0.32
EU-data US-dis US-work US-scale	59.5%	36.5%	25.0%	15.1%	6.3%	0.34
EU-data US-dis US-work EU-scale	61.8%	40.5%	28.6%	18.0%	7.8%	0.33
EU-data US-dis EU-work US-scale	34.4%	22.3%	16.7%	11.5%	6.1%	0.20
EU-data US-dis EU-work EU-scale	35.3%	24.4%	18.8%	13.4%	7.4%	0.19
EU-data EU-dis US-work US-scale	59.8%	35.3%	24.3%	15.0%	6.7%	0.33
EU-data EU-dis US-work EU-scale	61.8%	39.3%	27.8%	17.7%	8.2%	0.33
EU-data EU-dis EU-work US-scale	33.6%	20.7%	15.4%	10.8%	6.1%	0.20
EU	34.5%	22.9%	17.4%	12.4%	7.3%	0.19

A comparison across the rows of Table 8 (or the columns of Figure 5) shows that a major source of differences lies in the parameters of the employment equation. Whenever we simulate labor market outcomes using the US employment parameters, we find considerably higher employment rates than when we use EU parameters. This is particularly true for the less severe

work disability categories, in line with the larger effect of work disability on employment in the US than in Europe. Once we move to extreme work disabilities the differences become small; with an extreme work disability neither Americans nor Europeans are likely to work.

Figure 5: Percent working by disability category
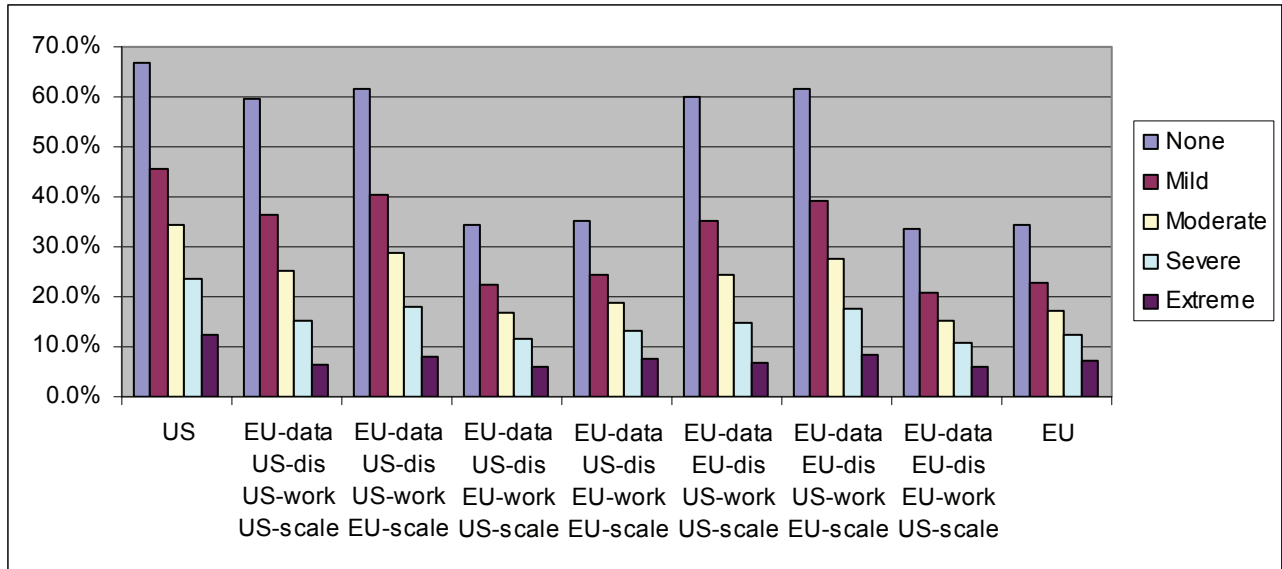
Figure 6 shows total employment rates in the US and the EU. Consistent with Table 8 and Figure 5, we see that employment is about twice as large when a US employment equation is used than when we use the EU parameters in the employment equation. Whether the EU or the US disability parameters are used has very little impact.

Finally, we return to the issue of justification bias. The last column of Table 8 shows simulated correlations between the binary outcomes employment (employment=1 and non-employment=0) and not having a work limiting health condition (none=1, mild or worse=0). Other things equal, one would expect a stronger correlation between work and self-reported

disability if US scales are used, because of justification bias, which is significant in the US and absent in the EU. We see a small effect of justification bias. For instance, comparing the second and third row we see the correlation decrease from .34 to .33. Other comparisons show similar small effects. However using the US employment parameters rather than the EU ones has a much bigger effect. This confirms once again the much weaker relation between disability and employment in the EU than in the US.

Figure 6: Employment rates
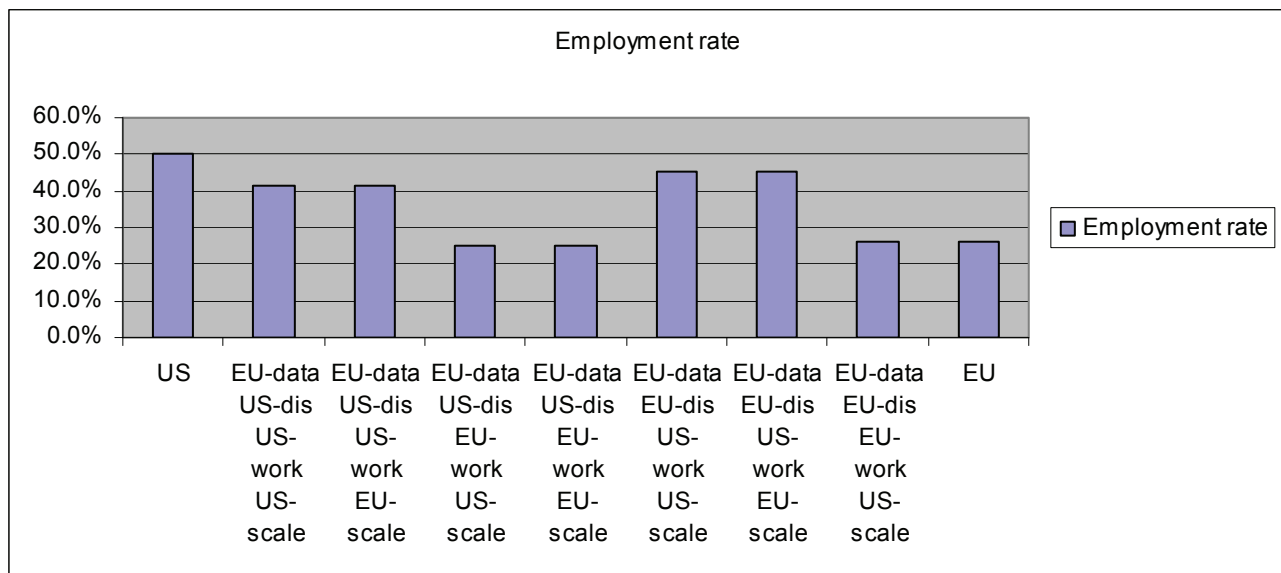
8. Conclusions

We have provided several pieces of evidence on differences in work disability between older Europeans and Americans. The descriptive vignette analyses suggest that opinions about what constitutes a health related work limitation are related to the generosity of earnings replacement schemes and employment protection in different countries. We also find that the differences in self-reported disabilities across European countries fall substantially when we use

common reporting scales (i.e. if we use vignette corrections to make scales comparable). Consistent with earlier research, we find that Americans are less likely to call themselves work disabled than Europeans.

The extended model with employment included, suggests that justification bias plays a role in the US, but not in the EU. Americans use health as a justification for not working, whereas Europeans don't feel the need to do so. The effect of health limitations on employment is about twice as large in the US than in the EU. Our simulations suggest that this mainly reflects the fact that in the EU even healthy individuals are less likely to be working than in the US.

References

Böckerman, P., and P. Ilmakunnas (2009), "Unemployment and self-assessed health: evidence from panel data," *Health Economics*, 18(2), 161-179.

Boersch-Supan, A. (2007), "Work Disability, Health, and Incentive Effects," Working Paper, MEA, University of Mannheim.

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

Eurostat. (2001),.*Disability and Social Participation in Europe*, Luxembourg: Office for Official Publications of the European Communities.

Croda, E. and J. Skinner (2009). "Disability Insurance and Health in Europe and the US. April," Paper presented at Netspar 2009 Annual Conference, The Hague, http://www.netspar.nl/events/2009/annual/paperskinner.pdf.

Heeringa S.G., and J. Connor (1995), "Technical Description of the Health and Retirement Study Sample Design," Online version at www.hrsonline.isr.umich.edu; originally published as *HRS/AHEAD Documentation Report* DR-002.

Jones, M.K. (2007), "How accurate are estimates of the impact of disability on labour market participation? Evidence from the Health Survey for England," Swansea University, School of Business, Working Paper.

Kapteyn, A., J.P. Smith, and A. van Soest (2007), "Vignettes and self-reported work disability in the US and the Netherlands," *American Economic Review*, 97(1), 461-473.

Kapteyn, A., J.P. Smith, A. van Soest, and J. Banks (2009), "Labor Market Status and Transitions during the Pre-Retirement Years: Learning from International Differences," forthcoming in D. Wise (ed.), *Research Findings in the Economics of Aging*, University of Chicago Press.

King, G., C. Murray, J. Salomon, and A.Tandon (2004),"Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research," *American Political Science Review,* 98(1), 567-583.

Kreider, B. and J.V. Pepper (2007), "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of the American Statistical Association*, 102(478), 432-441.

OECD (2004), *OECD Employment Outlook,* OECD, Paris.

Van Soest, A., L. Delaney, C. Harmon, A. Kapteyn, and J.P. Smith (2007), "Validating the Use of Vignettes for Subjective Threshold Scales," RAND, Labor and Population Working Paper, http://www.rand.org/pubs/working_papers/WR501/

APPENDIX A. Tables

Table A.1: Responses to Pain Work Disability Vignettes

pain 1	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	1.96	0.65	0.41	4.03	2.12	1.51	1.56	1.28	2.05	1.53
mild	8.91	4.8	11.65	9.33	4.48	4.5	9.14	6.41	8.02	8.54
moderate	35.08	22.56	26	30.27	28.55	13.98	40.68	31.64	31.25	34.99
severe	47.77	62.2	41.12	43.47	36.49	53.16	43.83	42	47.45	48.32
extreme	6.29	9.79	20.82	12.89	28.35	26.85	4.78	18.67	11.24	6.63
Total	100	100	100	100	100	100	100	100	100	100

pain 2	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	2.28	1.81	3.4	6.58	2.63	0.94	5.29	2.9	3.72	24.97
mild	26.77	16.38	28.18	31.47	56.23	9.48	33.31	44.83	29.04	52.9
moderate	54.94	44.95	48.07	48.06	29.92	35.53	52.75	41.81	48.51	18.78
severe	13.85	34.78	17.22	11.73	10.72	47.59	7.67	9.59	16.71	2.25
extreme	2.15	2.09	3.13	2.16	0.5	6.46	0.98	0.87	2.02	1.11
Total	100	100	100	100	100	100	100	100	100	100

pain 3	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	1.22	0	0.41	2.62	1.47	0.45	1.15	1.35	1.28	1.16
mild	6.29	1.94	7.39	11.78	3.46	10.93	4.7	5.86	6.62	3.72
moderate	22.06	25.41	23.77	33.73	35.78	37.14	25.7	32.88	27.61	21.06
severe	57.57	63.15	48.03	34.96	36.75	43.78	61.77	47.86	51.59	50.43
extreme	12.85	9.5	20.41	16.91	22.53	7.69	6.68	12.05	12.9	23.62
Total	100	100	100	100	100	100	100	100	100	100

Table A.2: Responses to Affect Disability Vignettes

Affect 1	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	3.78	1.12	3.78	6.27	3.94	5.32	5.28	5.54	4.3	22.12
mild	20.07	22.62	16.77	32.91	25.37	28.69	21.59	29.44	24.4	31.15
moderate	48.64	45.91	43.7	43.25	46.54	49.47	50.15	41.55	46.68	32.21
severe	24.78	28.72	25.68	15.33	20.26	15.49	21.7	22.56	22.12	13.26
extreme	2.71	1.62	10.07	2.25	3.89	1.03	1.28	0.91	2.5	1.26
Total	100	100	100	100	100	100	100	100	100	100

Affect 2	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	3.38	1.88	2.93	5.93	2.41	1.41	4.84	3.66	3.82	18.33
mild	24.81	19.08	19.48	29.67	16.26	13.19	21.35	18.73	23.05	37.12
moderate	49.11	48.43	44.4	47.51	50.92	32.57	49.35	51.87	48.04	32.35
severe	21.62	27.21	27.15	14.86	25.29	45.02	22.52	22.21	22.48	10.74
extreme	1.08	3.4	6.04	2.03	5.13	7.8	1.94	3.52	2.6	1.46
Total	100	100	100	100	100	100	100	100	100	100

Affect 3	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	8.58	3.41	10.81	8.87	5.9	0.91	13.09	10.87	8.33	39.54
mild	46.2	35.9	33.06	48.36	68.9	8.94	39.77	56.78	43.66	41.02
moderate	40.69	43.28	40.22	32.11	18.69	32.32	41.38	27.99	37.23	15.86
severe	4.53	16.98	11.87	8.96	6.51	47.38	5.02	4.15	9.61	2.71
extreme	0	0.44	4.04	1.7	0	10.45	0.74	0.22	1.16	0.87
Total	100	100	100	100	100	100	100	100	100	100

Table A.3: Responses to Heart Disease Work Disability Vignettes

CVD1	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	3.1	0.63	1.41	4.08	2.47	0.15	4.11	3.3	2.88	7.84
mild	12.78	5.1	13.44	19.61	31.1	1.52	11.88	18.86	13.85	35.41
moderate	46.57	30.52	33.26	46.33	38.96	9.78	36.27	45.97	39.82	37.76
severe	34.47	57.95	40.42	23.96	22.37	60.56	43.69	26.32	37.6	16.12
extreme	3.08	5.81	11.46	6.02	5.09	27.99	4.06	5.55	5.85	2.87
Total	100	100	100	100	100	100	100	100	100	100

CVD2	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	3.31	0.3	0.6	4.47	3.76	1.95	3.62	2.46	2.98	8.16
mild	11.27	4.09	5.16	20.73	26.7	24.3	9.95	18.89	13.47	20.46
moderate	43.46	29.74	17.2	38.65	42.9	47.68	39.5	45.51	38.5	40.88
severe	39.15	54.3	49.3	24.12	21.15	22.22	41.95	30.44	37.01	26.93
extreme	2.45	11.57	27.74	8.31	5.49	3.85	4.98	2.7	7.08	3.57
missing	0.36	0	0	3.72	0	0	0	0	0.96	100
Total	100	100	100	100	100	100	100	100	100	

CVD3	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	US
none	3.99	0.45	1.86	5.31	3.04	0.15	2.14	3.83	3.12	3.1
mild	3.2	3.39	3.25	10.75	15.15	2.12	4.77	6.93	5.99	7.55
moderate	21.55	14.35	8.96	21.55	35.13	11.4	17.34	29.66	19.8	24.99
severe	53.39	57.53	33.63	36.62	32.8	54.02	55.52	42.04	48.11	48.61
extreme	17.87	24.28	52.29	25.77	13.89	32.32	20.23	17.54	22.99	15.74
Total	100	100	100	100	100	100	100	100	100	100

Table A.4: Estimates Model Section 3: Work Disability (Eq. (1))

	Model without DIF		Model with DIF	
	coeff.	st. error	coeff.	st. error
Constant	-0.135	0.107	-0.457*	0.131
Female	-0.103*	0.043	-0.020	0.049
Married/LT	-0.179*	0.045	-0.136*	0.052
Educyrs	-0.046*	0.007	-0.043*	0.008
Heart prob	0.487*	0.048	0.437*	0.054
Lung dis	0.421*	0.068	0.391*	0.077
High blood	0.145*	0.044	0.114*	0.049
Diabetes	0.325*	0.052	0.241*	0.061
Pain	0.431*	0.045	0.384*	0.05
Arthritis	0.352*	0.044	0.342*	0.049
Cancer	0.142*	0.061	0.185*	0.069
Cesd score	0.171*	0.011	0.139*	0.013
Obese	0.163*	0.047	0.130*	0.053
Age 58-64	0.154*	0.058	0.141*	0.065
Age 65-71	0.130*	0.056	0.112+	0.063
Age 72+	0.461*	0.064	0.372*	0.072
Female*EU	-0.014	0.058	-0.090	0.062
Marr/LT*EU	0.113+	0.063	0.084	0.068
Educyrs*EU	0.041*	0.007	0.036*	0.008
Heart*EU	-0.032	0.070	0.013	0.075
Lung*EU	-0.139	0.090	-0.044	0.097
Highbl*EU	-0.068	0.059	-0.047	0.063
Diabetes*EU	-0.115	0.079	-0.011	0.087
Pain*EU	0.005	0.060	0.067	0.064
Arthr*EU	-0.019	0.065	0.008	0.069
Cancer*EU	0.260*	0.098	0.235*	0.102
Cesd*EU	0.114*	0.020	0.148*	0.021
Obese*EU	0.008	0.067	0.035	0.072
Age58-64*EU	-0.015	0.077	-0.002	0.082
Age65-71*EU	0.132+	0.077	0.170*	0.083
Age72+*EU	0.012	0.083	0.116	0.090
Germany	-0.497*	0.139	-0.913*	0.153
Sweden	-0.600*	0.135	-1.130*	0.152
Netherland	-0.609*	0.137	-0.993*	0.151
Spain	-0.642*	0.134	-1.279*	0.149
Italy	-0.790*	0.137	-1.112*	0.152
France	-0.840*	0.130	-1.244*	0.145
Greece	-1.254*	0.133	-1.767*	0.149
Belgium	-0.535*	0.135	-0.891*	0.150

*,+: significant at two-sided 5% and 10% level, respectively

Table A.5: Estimates Model Section 3: Thresholds (Eq. (3))

	Threshold 1		log(threshold 2 - threshold 1)		log(threshold 3 - threshold 2)		log(threshold 4 - threshold 3)	
	parameter	standard error	parameter	standard error	parameter	standard error	parameter	standard error
US								
Constant	0	0	-0.726*	0.048	-0.836*	0.049	-0.584*	0.052
Female	0.070*	0.016	0.040*	0.018	0.012	0.019	-0.072*	0.020
Married/LT	-0.011	0.017	0.071*	0.020	0.062*	0.020	0.043*	0.021
Educyrs	-0.012*	0.003	0.017*	0.003	0.019*	0.003	0.024*	0.003
Heart prob	-0.023	0.020	0.017	0.022	-0.008	0.022	0.004	0.023
Lung dis	-0.004	0.030	-0.030	0.032	0.026	0.031	0.001	0.035
High blood	-0.014	0.017	-0.008	0.018	-0.003	0.019	-0.011	0.020
Diabetes	-0.068*	0.019	-0.002	0.024	0.027	0.024	-0.055*	0.026
Pain	-0.053*	0.018	0.020	0.019	0.036+	0.019	0.002	0.022
Arthritis	-0.017	0.017	0.021	0.019	0.000	0.020	0.030	0.021
Cancer	0.044+	0.025	0.014	0.026	0.014	0.028	0.051+	0.031
Cesd score	-0.013*	0.005	-0.008	0.005	-0.014*	0.005	-0.016*	0.005
Obese	-0.008	0.017	-0.021	0.020	-0.022	0.021	-0.018	0.022
Age 58-64	-0.021	0.024	0.016	0.025	0.019	0.027	0.050+	0.028
Age 65-71	-0.037+	0.021	-0.010	0.024	0.039	0.025	0.051+	0.027
Age 72+	-0.108*	0.024	0.048+	0.028	0.126*	0.028	0.080*	0.030
Interaction EU								
Female	-0.077*	0.022	0.009	0.025	-0.039+	0.022	0.047+	0.024
Married/LT	0.021	0.024	-0.061*	0.028	-0.060*	0.025	-0.042	0.026
Educyrs	0.007*	0.003	-0.013*	0.003	-0.018*	0.003	-0.022*	0.003
Heart prob	0.012	0.030	-0.041	0.035	0.005	0.029	-0.009	0.031
Lung dis	0.052	0.039	0.045	0.044	-0.047	0.037	-0.017	0.043
High blood	-0.013	0.023	0.025	0.026	0.046*	0.023	-0.012	0.025
Diabetes	0.156*	0.029	-0.098*	0.038	-0.065*	0.032	0.047	0.034
Pain	0.014	0.025	0.013	0.026	0.023	0.023	-0.021	0.026
Arthritis	0.023	0.027	-0.023	0.029	-0.039	0.025	-0.026	0.028
Cancer	-0.085*	0.038	0.043	0.043	0.007	0.037	-0.004	0.043
Cesd score	-0.005	0.009	-0.001	0.010	0.026*	0.007	0.027*	0.008
Obese	0.061*	0.027	-0.055+	0.030	-0.014	0.027	-0.058*	0.028
Age 58-64	0.033	0.031	0.020	0.033	0.012	0.031	-0.001	0.034
Age 65-71	0.061*	0.029	0.048	0.034	-0.023	0.030	-0.016	0.033
Age 72+	0.093*	0.033	-0.012	0.038	-0.066*	0.033	0.017	0.035
Country Dummies								
Germany	-0.639*	0.054	0.338*	0.062	0.586*	0.055	0.560*	0.059
Sweden	-0.657*	0.062	0.180*	0.069	0.285*	0.057	0.419*	0.058
Netherland	-0.659*	0.051	0.665*	0.058	0.478*	0.055	0.134*	0.059
Spain	-0.915*	0.059	0.447*	0.063	0.550*	0.055	0.637*	0.056
Italy	-0.527*	0.051	0.441*	0.058	0.558*	0.054	0.160*	0.059
France	-0.575*	0.051	0.289*	0.058	0.589*	0.053	0.551*	0.057
Greece	-0.662*	0.057	0.238*	0.060	0.395*	0.054	0.217*	0.057
Belgium	-0.630*	0.054	0.523*	0.059	0.521*	0.054	0.313*	0.058

*,+: significant at two-sided 5% and 10% level, respectively

Table A.6: Estimates Model Section 3: Other Parameters & Log Likelihood

	Model without DIF		Model with DIF	
	parameter	standard error	parameter	standard error
Vignette dummies				
d vig pain 1	1.621*	0.024	1.003*	0.044
d vig pain 2	0.747*	0.015	0.202*	0.042
d vig pain 3	1.779*	0.025	1.145*	0.045
d vig aff 1	0.928*	0.017	0.372*	0.042
d vig aff 2	0.982*	0.018	0.424*	0.043
d vig aff 3	0.563*	0.014	0.019	0.042
d vig cvd 1	1.217*	0.019	0.641*	0.043
d vig cvd 2	1.309*	0.020	0.720*	0.043
d vig cvd 3	1.798*	0.025	1.167*	0.045
Standard deviations of errors / unobserved heterogeneity				
sig selfreport	1	(normalization)	1	(normalization)
sig threshold			0.423*	0.007
sig vignette	0.728*	0.009	0.513*	0.007
Threshold parameters				
const thrh	0	(normalization)	See Table A.5	
const thrh	-0.299*	0.013		
const thrh	-0.290*	0.013		
const thrh	-0.109*	0.013		
Log likelihood	-96584.35		-86369.50	

*= significant at two-sided 5% level.

Table A.7: Estimates Model Section 6: Work Disability Equation (Eq. (1))

	Model without DIF		Model with DIF	
	coeff.	St. error	coeff.	st. error
constant	-0.126	0.107	-0.189	0.118
Female	-0.106*	0.043	-0.017	0.045
Married/LT	-0.175*	0.045	-0.118*	0.047
Educyrs	-0.046*	0.007	-0.039*	0.007
Heart prob	0.484*	0.048	0.463*	0.050
Lung dis	0.423*	0.068	0.421*	0.071
High blood	0.139*	0.044	0.112*	0.045
Diabetes	0.321*	0.052	0.256*	0.056
Pain	0.426*	0.045	0.411*	0.047
Arthritis	0.354*	0.044	0.364*	0.046
Cancer	0.142*	0.061	0.206*	0.064
Cesd score	0.170*	0.011	0.148*	0.011
Obese	0.167*	0.047	0.137*	0.049
Age 58-64	0.157*	0.058	0.166*	0.061
Age 65-71	0.138*	0.056	0.135*	0.059
Age 72+	0.463*	0.064	0.426*	0.068
Female*EU	-0.007	0.058	-0.075	0.060
Marr/LT*EU	0.110+	0.063	0.078	0.065
Educyrs*EU	0.042*	0.007	0.034*	0.007
Heart*EU	-0.027	0.070	-0.033	0.073
Lung*EU	-0.141	0.090	-0.098	0.094
Highbl*EU	-0.065	0.059	-0.028	0.061
Diabetes*EU	-0.110	0.079	-0.076	0.083
Pain*EU	0.006	0.060	0.048	0.062
Arthr*EU	-0.019	0.065	-0.038	0.067
Cancer*EU	0.262*	0.098	0.209*	0.102
Cesd*EU	0.114*	0.020	0.126*	0.020
Obese*EU	0.004	0.067	-0.020	0.069
Age58-64*EU	-0.018	0.077	0.004	0.079
Age65-71*EU	0.122	0.077	0.161*	0.080
Age72+*EU	0.010	0.083	0.087	0.087
Germany	-0.503*	0.139	-0.689*	0.145
Sweden	-0.612*	0.136	-1.066*	0.142
Netherland	-0.615*	0.137	-0.699*	0.142
Spain	-0.646*	0.134	-0.999*	0.140
Italy	-0.798*	0.137	-0.895*	0.142
France	-0.848*	0.13	-1.023*	0.136
Greece	-1.260*	0.133	-1.690*	0.139
Belgium	-0.542*	0.135	-0.654*	0.141

*: significant at two-sided 5% level; + significant at two-sided 10% level

Table A.8: Estimates Model Section 6: Thresholds Equation (Eq. (8); model with DIF)

	coeff.	st. error		coeff.	st. error
Work	0.097*	0.015	Work*EU	-0.104*	0.023
Female	0.096*	0.013	Female*EU	-0.082*	0.019
Married/LT	0.051*	0.013	Marr/LT*EU	-0.031	0.020
Educyrs	0.003	0.002	Educyrs*EU	-0.004+	0.002
Heart prob	-0.016	0.016	Heart*EU	-0.014	0.027
Lung dis	-0.002	0.022	Lung*EU	0.037	0.033
High blood	-0.024+	0.013	Highbl*EU	0.028	0.020
Diabetes	-0.065*	0.014	Diabetes*E	0.069*	0.025
Pain	-0.038*	0.015	Pain*EU	0.046*	0.022
Arthritis	-0.005	0.014	Arthr*EU	-0.008	0.024
Cancer	0.066*	0.020	Cancer*EU	-0.048	0.036
Cesd score	-0.022*	0.004	Cesd*EU	0.008	0.008
Obese	-0.039*	0.014	Obese*EU	0.014	0.023
Age 58-64	0.021	0.020	Age58-64*EU	0.034	0.028
Age 65-71	0.020	0.016	Age65-71*EU	0.039	0.027
Age 72+	0.021	0.020	Age72+*EU	0.026	0.030
			Germany	-0.187*	0.047
const thrh 1	0	0	Sweden	-0.452*	0.051
thr2 - thr1	0.722*	0.010	Netherland	-0.074+	0.043
thr3 - thr2	0.704*	0.010	Spain	-0.356*	0.049
thr4 - thr3	0.822*	0.012	Italy	-0.083*	0.042
			France	-0.147*	0.043
sigma u	0.426*	0.007	Greece	-0.417*	0.046
			Belgium	-0.114*	0.046

*: significant at two-sided 5% level; + significant at two-sided 10% level

Table A.9: Estimates Model Section 6: Vignette Equation (Eq. (5); model with DIF)

	coeff.	st. error
dvigpain1	1.453*	0.038
dvigpain2	0.631*	0.034
dvigpain3	1.598*	0.039
dvigaff1	0.803*	0.034
dvigaff2	0.853*	0.035
dvigaff3	0.455*	0.034
dvigcvd1	1.075*	0.036
dvigcvd2	1.161*	0.036
dvigcvd3	1.615*	0.039
Sigmav	0.527*	0.007

*: significant at two-sided 5% level; + significant at two-sided 10% level

Table A.10: Estimates Model Section 6: Employment Equation (Eq. (10))

	Model without DIF		Model with DIF	
	coeff.	st. error	coeff.	st. error
Disability	-0.516*	0.038	-0.464*	0.036
Disability*EU	0.319*	0.049	0.272*	0.049
Constant	0.473*	0.151	0.440*	0.135
Female	-0.289*	0.059	-0.246*	0.057
Married/LT	-0.069	0.065	-0.034	0.062
Educyrs	0.030*	0.009	0.034*	0.009
Heart prob	0.015	0.076	-0.015	0.074
Lung dis	-0.102	0.109	-0.125	0.107
High blood	0.016	0.059	-0.001	0.057
Diabetes	-0.189*	0.079	-0.228*	0.076
Pain	0.131*	0.064	0.102+	0.061
Arthritis	0.041	0.062	0.031	0.060
Cancer	-0.023	0.084	0.004	0.085
Cesd score	-0.035*	0.017	-0.051*	0.016
Obese	0.223*	0.064	0.204*	0.059
Age 58-64	-0.633*	0.079	-0.620*	0.078
Age 65-71	-1.285*	0.077	-1.268*	0.073
Age 72+	-1.815*	0.101	-1.813*	0.100
Female*EU	-0.282*	0.081	-0.319*	0.079
Marr/LT*EU	-0.096	0.094	-0.125	0.092
Educyrs*EU	-0.015	0.010	-0.020*	0.009
Heart*EU	-0.021	0.128	-0.002	0.127
Lung*EU	-0.032	0.161	0.002	0.160
Highbl*EU	0.034	0.085	0.051	0.084
Diabetes*EU	0.122	0.128	0.152	0.127
Pain*EU	-0.058	0.086	-0.025	0.085
Arthr*EU	-0.066	0.101	-0.058	0.100
Cancer*EU	0.058	0.155	0.034	0.155
Cesd*EU	-0.039	0.032	-0.026	0.031
Obese*EU	-0.296*	0.097	-0.284*	0.094
Age58-64*EU	-0.399*	0.098	-0.406*	0.097
Age65-71*E	-1.288*	0.133	-1.296*	0.130
Age72+*EU	-1.407*	0.211	-1.402*	0.211
Germany	0.310	0.196	0.291	0.184
Sweden	0.852*	0.206	0.784*	0.196
Netherland	0.239	0.195	0.241	0.183
Spain	0.252	0.192	0.208	0.181
Italy	-0.129	0.197	-0.126	0.186
France	0.310+	0.189	0.300+	0.177
Greece	0.215	0.191	0.157	0.183
Belgium	0.044	0.191	0.041	0.178
Log-likelihood (of complete model)	-99650.54		-90792.8	

*: significant at two-sided 5% level; + significant at two-sided 10% level.

Table A.11: Estimates Model Section 6: Employment equation without objective health conditions (Eq. (10)) & Log Likelihood

	Model without DIF		Model with DIF	
	coeff.	st.er.	coeff.	St.er.
Health	-0.506*	0.032	-0.477*	0.031
Health*EU	0.282*	0.041	0.257*	0.040
Constant	0.505*	0.142	0.399*	0.141
Female	-0.280*	0.057	-0.239*	0.057
Married/LT	-0.038	0.064	0.000	0.063
Educyrs	0.030*	0.009	0.036*	0.009
Age 58-64	-0.627*	0.075	-0.614*	0.075
Age 65-71	-1.280*	0.070	-1.261*	0.070
Age 72+	-1.830*	0.097	-1.825*	0.096
Female*EU	-0.313*	0.079	-0.347*	0.078
Marr/LT*EU	-0.103	0.093	-0.134	0.088
Educyrs*EU	-0.015+	0.009	-0.021*	0.009
Age58-64*EU	-0.391*	0.094	-0.398*	0.092
Age65-71*EU	-1.277*	0.125	-1.287*	0.125
Age72+*EU	-1.366*	0.205	-1.365*	0.204
Germany	0.218	0.182	0.260	0.181
Sweden	0.760*	0.194	0.748*	0.193
Netherland	0.132	0.183	0.199	0.182
Spain	0.129	0.180	0.142	0.179
Italy	-0.255	0.183	-0.188	0.181
France	0.175	0.174	0.226	0.172
Greece	0.084	0.174	0.080	0.165
Belgium	-0.067	0.178	-0.006	0.177
log likelihood (of complete model)	-99670.88		-90815.5	

*: significant at two-sided 5% level; + significant at two-sided 10% level.

49

APPENDIX B. The SHARE and HRS Work Disability Vignettes and Self-Reports

SHARE

Work limiting disabilities
Three domains: affect, pain, cardio-vascular disease.
One overall self-report on work-limiting disabilities.
Three vignettes per domain.

Response categories (for all questions; same as for health domains):
   1. None; 2. Mild; 3. Moderate; 4. Severe; 5. Extreme/Cannot Do

Self report:
Do you have any impairment or health problem that limits the kind or amount of work you can do?

Introduction to vignettes:
We would now like to give you a number of examples of persons with some health problems. We would like you to indicate the extent to which you think these people would be limited in the kind or amount of work they can do. In terms of their age, their education, and their work histories, you should imagine that these men or women are similar to yourself. Other than the conditions explicitly mentioned, you should imagine the individual is in reasonably good health.
General form of vignette questions:
How much was [name of person] limited in the kind or amount of work he/she could do?

Affect Vignettes:
1. [Eva] feels worried all the time. She gets depressed once a week at work for a couple of days in a row, thinking about what could go wrong and that her boss will disapprove of her condition. But she is able to come out of this mood if she concentrates on something else.

2. [Tamara] has mood swings on the job. When she gets depressed, everything she does at work is an effort for her and she no longer enjoys her usual activities at work. These mood swings are not predictable and occur two or three times during a month.

3. [Henriette] generally enjoys her work. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day-to-day activities on the job.

Pain Vignettes:
1. [Yvonne] has almost constant pain in her back and this sometimes prevents her from doing her work.

2. [Catherine] suffers from back pain that causes stiffness in her back especially at work but is relieved with low doses of medication. She does not have any pains other than this generalized discomfort.

3. [Mark] has pain in his back and legs, and the pain is present almost all the time. It gets worse while he is working. Although medication helps, he feels uncomfortable when moving around, holding and lifting things at work

CVD Vignettes:
1. [Norbert] has had heart problems in the past and he has been told to watch his cholesterol level. Sometimes if he feels stressed at work he feels pain in his chest and occasionally in his arms.

2. [Tom] has been diagnosed with high blood pressure. His blood pressure goes up quickly if he feels under stress. Tom does not exercise much and is overweight.

3. [Dan] has undergone triple bypass heart surgery. He is a heavy smoker and still experiences severe chest pain sometimes.

Randomization

All vignette questions are asked to all respondents. The only randomization concerns the names of the vignette persons and the order of the vignette questions. The SHARE questionnaire had two versions. Each has alternating male and female names on the vignettes, but where version 1 has a male name, version 2 has a female name, etc. Version 1 has the order of the questions as given above. Version 2 reverses things: first work disability, then health; domains in reverse order; vignettes in each domain in reverse order. The self-report always comes before the vignettes.

HRS

Although the vignettes and self-reports in HRS are very similar to those in SHARE, there are a few noticeable differences.

Self-report in 2004 leave-behind questionnaire:
To what extent are you limited in the kind or amount of work you can do because of an impairment or health problem?

Not at all limited, Mildly limited, Moderately limited, Severely limited, Cannot do any work

Vignettes
HRS fielded the same vignettes as used in Kapteyn et al. (2007), with some slight changes in formulation. The SHARE vignettes are a subset of these. Comparing the vignettes that are common between HRS and SHARE, we find one difference. SHARE vignette CVD-2 above in the HRS version reads:

Diane has been diagnosed with high blood pressure. Her blood pressure goes up quickly if she feels under stress. Diane does not exercise much and is overweight. Life can sometimes be hectic for her. She does not get along with her boss very well.