This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Explorations in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90337-0
ISBN13: 978-0-226-90337-8

Volume URL: http://www.nber.org/books/wise09-2

Conference Date: May 2009

Publication Date: March 2011

Chapter Title: Work Disability, Work, and Justification Bias in Europe and the United States

Chapter Authors: Arie Kapteyn, James P. Smith, Arthur van Soest

Chapter URL: http://www.nber.org/chapters/c11946

Chapter pages in book: (269 - 312)

9

Work Disability, Work, and Justification Bias in Europe and the United States

Arie Kapteyn, James P. Smith, and Arthur van Soest

9.1 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 United States are also found in self-reports of work-limiting disabilities. Table 9.1, taken from Kapteyn et al. (2009), illustrates the point.[1] We see considerable variation in both DI expenditures as a percentage of gross domestic product (GDP) and the percentage of males between forty and sixty-five reporting some form

Arie Kapteyn is a senior economist and director of the Labor and Population program at RAND Corporation. James P. Smith is a senior economist and holds the Distinguished Chair in Labor Markets and Demographic Studies at RAND Corporation. Arthur van Soest is professor of econometrics at Tilburg University, a research fellow at Netspar, and an affiliated researcher at RAND Corporation.

This research was supported by grants from the National Institute on Aging to RAND and the Social Security Administration through MRRC. It uses data from SHARE Wave 1, 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) and from HRS, funded by NIA. We are grateful to Angus Deaton for useful comments on an earlier version.

1. Self-reports for the European countries are taken from the European Community Household Panel (ECHP). For the United States they are taken from the Panel Study of Income Dynamics (PSID). 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?"

Table 9.1 **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	4.0	24.5
Belgium	2.2	14.3
France	1.7	20.5
UK	2.2	13.1
Ireland	1.3	15.7
Italy	2.0	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

of work disability. Remarkably, the two columns are only weakly correlated: the correlation is 0.20.

Croda and Skinner (2009), using data from Survey of Health, Aging, and Retirement in Europe (SHARE) and Health and Retirement Study (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 chapter 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 United States 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 chapter is organized as follows. The second section describes the data collection efforts in HRS and SHARE that make this research possible. Section 9.3 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 United States respond to the same vignette scenarios. Section 9.5 presents the empirical results and their implications for interpreting observed differences in self-reported work disability. Section 9.6 extends the model developed in section 9.3 to include employment and justification bias. Section 9.7 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.

9.2 Data Sources

For the United States, we use the 2004 wave of the Health and Retirement Study (HRS), a biannual panel with a representative sample of the U.S. population over fifty. It has been conducted by the University of Michigan since 1992. Information collected includes physical and mental health, socioeconomic status (including measures of labor market status, income, education and wealth), social support, and so forth. The surveys use a mixture of modes with most new interviews conducted face-to-face and most reinterviews by phone. Wave-specific overall response rates for the HRS have improved from 81.7 percent in 1992 to close to 90 percent at later waves, specifically 87.8 percent 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 seventy-five years of age.

The SHARE is a large-scale project that aims to collect interdisciplinary longitudinal data on European citizens age fifty 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, socioeconomic status including measures of income, education and wealth, and social support. The first wave of the main survey was fielded in 2004, with between 1,000 and 4,000 individuals in each country, adding up to about 29,000 individuals.

While containing many unique features, in order to facilitate additional comparisons with the United States and England, SHARE was purposely modeled after the HRS and the English Longitudinal Survey of Aging

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

(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, Smith, and van Soest (2007).

The work disability vignettes deal with work-limiting health problems in the domains of pain, depression, and cardiovascular 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

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 before. 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

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.

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 chapter 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."

scales. Moreover, the vignettes will also be used for comparisons of different socioeconomic groups.

9.3 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 9.1, 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 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

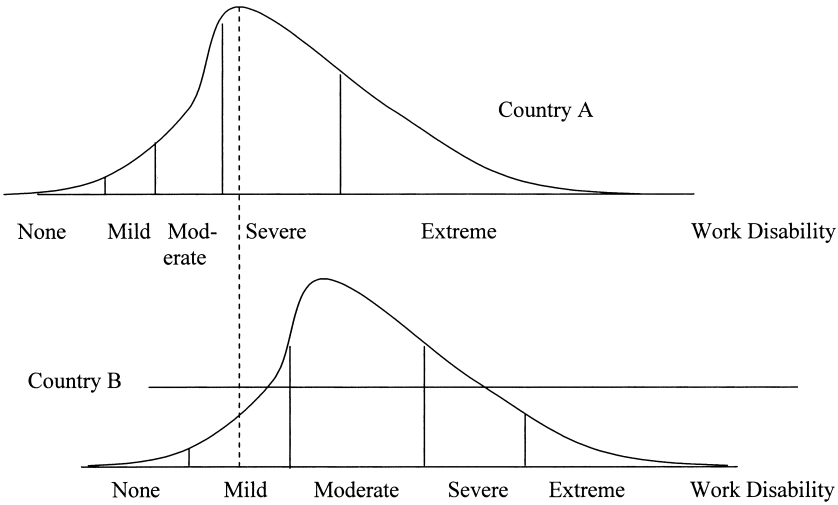

Fig. 9.1 Comparing self-reported health-related work limitations

disability distribution. Correcting for the differences in the response scales (DIF, or "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.

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 socioeconomic status (SES) health gradient reflects different health thresholds by SES rather than true health differences. Vignettes offer the potential for systematic testing of these hypotheses.

9.3.1 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, Smith, and van Soest 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 $l = 1, \ldots, L$.

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

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

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

The thresholds τ_j^i between the categories are given by

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

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

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 9.1.

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

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

7. Kapteyn, Smith, and van Soest (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.

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 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.

9.4 Responses to Vignettes in Western Europe and the United States

Respondents in each of the eight European countries and the United States 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 conditions of a hypothetical person. The actual vignettes used are presented in appendix B.[10] Table 9A.1 in appendix A compares the responses for the pain domain, while tables 9A.2 and 9A.3 do the same for the affect and heart disease (CVD) domains, respectively. The numbering of vignettes corresponds to how vignettes are presented in

8. Allowing the vignette evaluations to depend upon gender of the vignette person (as was done in Kapteyn, Smith, and van Soest 2007) does not affect the results.

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.

appendix B. For instance, pain1 is the first pain vignette shown in appendix B, pain2 is the second pain vignette, and so forth. Although the health conditions of the persons described in the vignettes are supposed to be the same in all countries, tables 9A.1 through 9A.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 percent 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 percent of Americans consider this to be no or only a mild work disability, while the SHARE average is about 16 percent. 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 9.2 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 9.2 summarizes the vignette evaluations in two ways. First, the 5-point scale on work disability is collapsed into three groups, with "none and mild" combined into one group and "severe and 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 9.2, 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 9.2, we first discuss some correlations summarized in table 9.3. 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

Table 9.2 **Responses to work disability vignettes**

	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	U.S.	Total EU
Pain										
pain 1										
none, mild	10.9	5.5	12.1	13.4	6.6	6.0	10.7	7.7	10.1	10.1
severe, extreme	54.1	72.0	61.9	56.4	64.8	80.0	48.6	60.7	55.0	58.7
pain 2										
none, mild	29.0	18.2	31.6	38.1	58.9	10.4	38.6	47.7	77.9	32.8
severe, extreme	16.0	36.9	20.4	13.9	11.2	54.0	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	34.0	26.9	35.0	53.3	28.7
severe, extreme	27.5	30.3	35.7	17.6	24.2	16.5	23.0	23.5	14.5	24.6
Affect 2										
none, mild	28.2	21.0	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	52.0
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	16.0	22.2	43.3	16.7
severe, extreme	37.5	63.8	51.9	30.0	27.5	88.5	47.7	31.9	19.0	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	77.0	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 9.3 **Correlations between rankings**

	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

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 9.3 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 9.3. 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 9.3 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 9.2 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 United States. Comparing the last two columns, which present the United States 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 among the European countries depends to some extent on the specific domain chosen. For example, the Italians are quite

Table 9.4 National norms about work disability and employment protection

		OECD employment protection indicator	
	Rank	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	3.0	2.9
Germany	5.0	2.2	2.5
Greece	6.7	2.8	2.9
Sweden	6.7	2.2	2.6
Spain	8.0	3.1	3.1
U.S.	2.6	0.2	0.7
Correlation with rank	1.00	0.52	0.56

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.

It is of interest to relate national norms about work disability to institutional arrangements. Table 9.4 provides a very simple way of doing so. The first column reproduces the grand average ranks from table 9.2. Columns (2) and (3) show two indicators of employment protection published by the Organization for Economic Cooperation and Development (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, 65):

1. Difficulty of dismissal; that is, legislative provisions setting conditions under which a dismissal is "justified" or "fair."
2. Procedural inconveniences that the employer may face when starting the dismissal process.
3. 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 do not find employment protection very important. This is then reflected in laws with little protection. The United States would be a case in point.

9.5 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 fifty-eight, fifty-eight to sixty-four, sixty-five to seventy-one, seventy-two 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, and the Center for Epidemiologic Studies Depression Scale (CES-D) score—for the European countries, the Euro-D scale. The benchmark country is the United States. We include a full set of interactions with an EU dummy. Furthermore, we include country dummies for the separate EU countries. Table 9.5 lists the means of the explanatory variables and self-reported work disability by country.

Panel A 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 sixty-five to seventy-one-year-olds in Sweden. Most chronic conditions are much more prevalent in the United States 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 United States 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, U.S. respondents underemphasize work-related health impairments, so that their actual work-related health limitations may be larger than those suggested by the means in panel B. Simulations based upon the estimates of the econometric model will look at this in detail.[11]

Table 9A.4 in appendix A presents estimation results for the complete model, allowing for DIF. Table 9A.5 presents the work disability equation in a model without DIF. The latter model is a standard ordered probit explaining the self-reported work disability on a 5-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.

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

Table 9.5 **Sample means by country**

A Demographics

Country	female	marrlt	educyrs	age 58–64	age 65–71	age 72+
U.S.	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

B Health Conditions

Country	heart	lung	hbp	diabetes	pain	arthritis
U.S.	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
U.S.	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

Notes: female: dummy female; marrlt: dummy married or living together (benchmark: single, divorced, separated, or widowed); educyrs: years of education; age 58–64, age 65–71, age 72+: dummies age groups 58–64, 65–71, and 72 and older, respectively (benchmark: 50–57). All sample means are weighted. 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 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 depression on work disability are stronger in the European countries than in the United States. No such effects are found in the model that does not correct for DIF (table 9A.5). The country dummies in table 9A.4 demonstrate that, once we correct for response scale differences and control for health conditions, the Europeans are considerably less likely to be work disabled than the Americans (the United States is the reference category). The effects are similar, but smaller, in the model without DIF, in accordance with the fact that U.S. respondents less easily classify someone as work disabled (compare to table 9.2). 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 United States.

The significant estimates of the effect of covariates on the thresholds 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 9.2. 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 European Union than in the United States. Females are harder in the United States, but not in the European Union. Older respondents are softer in the United States, but not in the European Union.

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 (compare to equation [3]). Rather than extensively discussing these parameters, we turn to some simulations that are easier to interpret.

9.5.1 Simulation Results

Figure 9.2 provides a first impression of the effects of using different scales. The first column presents the distribution of self-reported work disability for the United States, using U.S. data and U.S. parameters. This simply predicts sample observations for the United States. 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 European Union and the United States. For instance, the percentage saying that they have no work disability is predicted to be about 51 percent in the United States and 49 percent in the European Union. On the other hand, the percentage

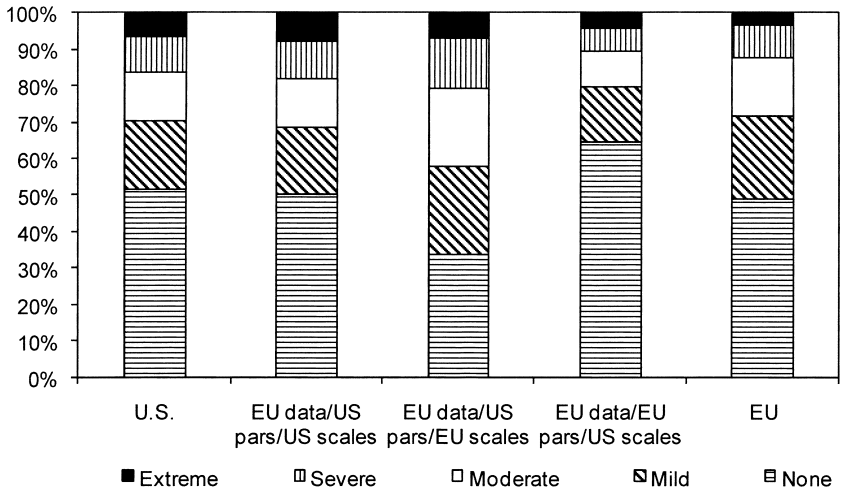

Fig. 9.2 Simulated work limitations according to different scenarios

of Americans with severe or extreme work disabilities is predicted to be about 16 percent in the United States and only 12 percent in the European Union. These seemingly modest differences are the result of two counteracting 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 9.2.

The second column uses EU data, but U.S. scales and U.S. 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 United States and the European countries.

The middle column is based on EU data, U.S. 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 U.S. 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 United States than in Europe. For instance, column (3) implies that 34 percent has no work limitation, while column (5) implies that 49 percent 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.

Table 9.6 **Work disability by country**

	None	Mild	Moderate	Severe	Extreme	None/ mild	Severe/ extreme
Germany							
U.S.	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, U.S. 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							
U.S.	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, U.S. 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							
U.S.	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, U.S. 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							
U.S.	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, U.S. 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							
U.S.	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, U.S. 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, U.S. 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							
U.S.	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, U.S. 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							
U.S.	51.4	19.1	13.2	9.9	6.5	70.4	16.4
EU, U.S. 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 we assign U.S. 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 percent to 64 percent, while the percentage with severe or extreme work limitations falls from 12 percent to about 10.5 percent. Again, this is a consequence of the fact that Americans less easily classify someone with a given impairment as work disabled.

Table 9.6 presents simulated work disability by country. The first row in each panel presents work disability for the United States; the second row presents work disability in the country, but using U.S. 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 U.S. 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 percent to 7.7 percent) and Spain (from 14.6 percent to 10.5 percent). In other countries the changes are much less dramatic and in some cases the adoption of the U.S. 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).

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 U.S. scales actually increases the difference in work disability between the U.S. 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 (U.S.) 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.

Next we consider a number of socioeconomic categories. Table 9.7 shows results by age. The scale effects appear to be fairly uniform by age. For instance, the effect of going from European scales to U.S. scales on the percentage severe/extreme varies from 1.2 in the fifty-eight to sixty-four category to 1.8 in the seventy-two-plus category. Similarly, the effect on the percentage none/mild varies from –7.5 in the fifty-eight to sixty-four category to –9.9 in the seventy-two-plus category.

9.6 Work Disability, Work, and Response Scales

We now extend the model presented in section 9.3 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 nonemployment on self-reported work disability, where respondents may exaggerate their work limitations to justify that they do not work. The em-

Table 9.7 **Work disability by age in the European Union and the United States**

	50–57			58–64			65–71			72+			All		
	U.S.	EU U.S. scale	EU	U.S.	EU U.S. scale	EU	U.S.	EU U.S. scale	EU	U.S.	EU U.S. scale	EU	U.S.	EU U.S. scale	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

pirical 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 5-point scale self-assessed work disability question while the answers to the work limitations of the vignette persons on the same 5-point scale are denoted by Y_{li}, where each respondent i evaluates L vignettes $l = 1, \ldots, L$.

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, etc.) 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 (compare equations [1] and [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) and (4) by:

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

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

This is a more parsimonious specification than the one used in equation (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 U.S. 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 nonemployed respondents.

The evaluations of vignettes $l = 1, \ldots, L$ are also modeled as before (compare equations [5] through [7]). Employment status is modeled using a probit-like equation:

$$E_i^* = X_i\varphi + \pi Y_i^* + \eta_i \tag{10}$$

$$\eta_i \sim N(0,1), \eta_i \text{ independent of } X_i, u_i, \varepsilon_{ri}, \varepsilon_{1i}, \ldots, \varepsilon_{Li}$$

$$E_i = 1 \text{ if } E_i^* > 0; E_i = 0 \text{ if } E_i^* \leq 0.$$

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).

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 9.3. 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 overidentified 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 9.3 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

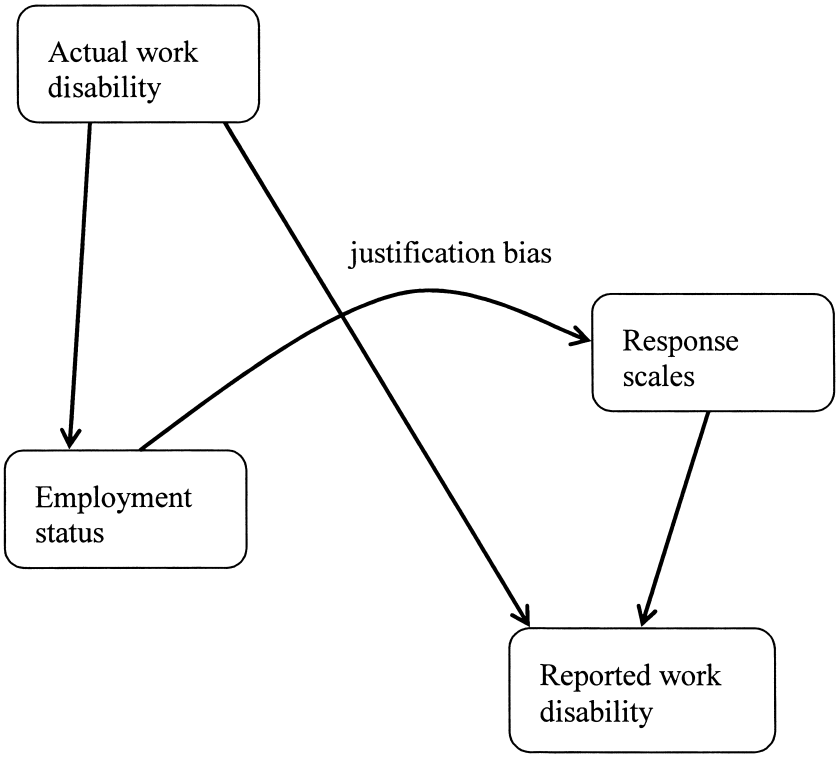

Fig. 9.3 Work, work disability, response scales, and justification bias

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 9.3, in particular vignette equivalence and response consistency. The first now also implies that workers and nonworkers 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 nonworkers 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.

9.7 Model Estimates for Work Disability, Work, and Threshold Equations

Tables 9A.6 through 9A.10 present the model estimates with and without DIF. Comparing table 9A.6 with the estimates for the disability equation in table 9A.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 9A.6, and which was completely insignificant in table 9A.4.

The results for the threshold equation given in table 9A.7 cannot be compared directly to table 9A.4, as the specification is much simpler. We will use simulations (shown later) to see if the simpler specification leads to very different outcomes. The country dummies (with the United States 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 9A.7 is the coefficient on the dummy work. The coefficient is significantly positive in the United States, while the interaction with the EU dummy makes clear that the effect of employment status on the thresholds is essentially zero in the European Union. In other words, we find significant justification bias in the United States and no evidence of justification bias in the European Union. It suggests that not working is much more accepted in the European Union (for whatever reason) than in the United States. 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 United States. It might mean that justification bias is present in Europe as well as the United States, but then it is still much bigger in the United States than in Europe.

The estimates of the vignette equation are presented in table 9A.8. The estimated coefficients on the vignette dummies are similar to those in table 9A.4, 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 9A.9. 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 CES-D score) and, in the United States, 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 United States 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 European Union than in the United States. Similarly, females are less likely to be employed than males and also this difference is stronger in the European Union than in the United States.

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 United States than in the European Union (–.464 in the United States and –.192 in the European Union). This result remains essentially unchanged when we specify the employment equations more parsimoniously by omitting the health conditions (table 9A.10). One possible interpretation of this is that individuals in the United States are more likely to work for pay and that health is one of the main impediments for doing so. In the European Union 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. Following, we will present simulations that shed more light on this (and other) interpretations.

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 9A.9 (–.192 and –.197, respectively), which corresponds to the finding that there is no evidence of justification bias for these countries. For the United States 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 nonemployed tend to overreport 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 9.6).

9.7.1 Simulation Results

Figure 9.4 is analogous to figure 9.2, but now based on the extended model, including an employment equation and allowing for justification bias. Compared to the model underlying figure 9.2, 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 9.4 are most comparable to those in figure 9.2, while the last two bars in figure 9.4 correspond to the last two bars in figure 9.2. Qualitatively the patterns appear to be similar. If we move from U.S. data to EU data (but retaining U.S. scales and parameters)—that is, from the first column to the second—we see again a slight shift in the direction of more work disability in the European Union than in the United States.

When we move from the second to the third column (EU data, U.S. parameters, EU scales) we see an increase in work disability, but the shift is not as dramatic as in figure 9.2. Finally, a comparison of the last two columns shows again that if we move from EU data and EU parameters, but U.S. scales to all data, parameters and scales European, then work disability increases substantially (but not as much as in figure 9.2). This illustrates once again the difference in response scales used in the EU countries and in the United States. The reason that the changes are less dramatic than in figure 9.2 is probably the fact that we have imposed that all threshold shifts are parallel to obtain a parsimonious model (see section 9.6).

Comparing columns (4) (EU-data, U.S.-disability equation, EU-work equation, U.S.-scale) and (8) (EU-data, EU-disability equation, EU-work equation, U.S.-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, U.S.-work equation, U.S.-scale) and (8) (EU-data, EU-disability

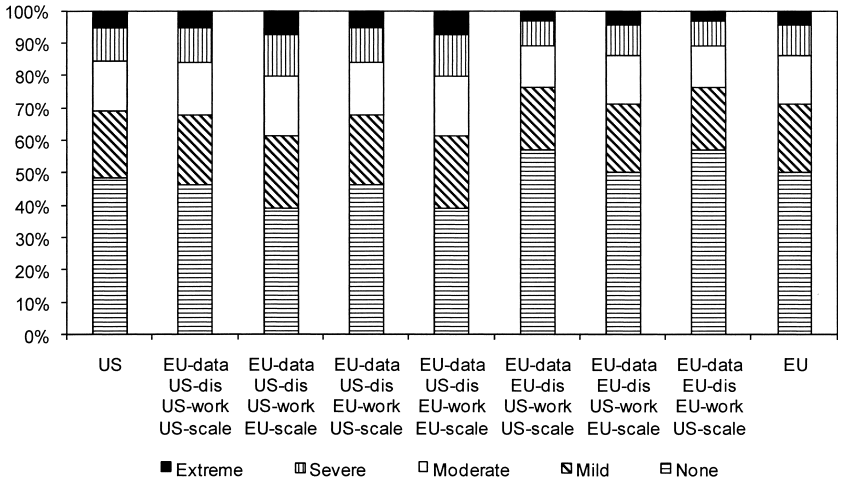

Fig. 9.4 Simulated work limitations; model including employment equation

equation, EU-work equation, U.S.-scale) is that in the former the U.S. employment parameters have been used, while in the latter 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 9.8 and figure 9.5 show the percentage of people working for each of the five levels of work disability (the last column of table 9.8 is discussed later). As could be expected on the basis of the raw data, a comparison of the first and last row of table 9.8 shows that Europeans work less than Americans for any work disability category. A comparison across the rows of table 9.8 (or the columns of figure 9.5) shows that a major source of differences lies in the parameters of the employment equation. Whenever we simulate labor market outcomes using the U.S. 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 United States 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 9.6 shows total employment rates in the United States and the European Union. Consistent with table 9.8 and figure 9.5, we see that employment is about twice as large when a U.S. employment equation is used than when we use the EU parameters in the employment equation. Whether the EU or the U.S. disability parameters are used has very little impact.

Finally, we return to the issue of justification bias. The last column of table

Table 9.8 **Percent working by disability category**

	None (%)	Mild (%)	Moderate (%)	Severe (%)	Extreme (%)	Correlations
U.S.	66.6	45.8	34.3	23.5	12.2	0.32
EU-data U.S.-dis U.S.-work U.S.-scale	59.5	36.5	25.0	15.1	6.3	0.34
EU-data U.S.-dis U.S.-work EU-scale	61.8	40.5	28.6	18.0	7.8	0.33
EU-data U.S.-dis EU-work U.S.-scale	34.4	22.3	16.7	11.5	6.1	0.20
EU-data U.S.-dis EU-work EU-scale	35.3	24.4	18.8	13.4	7.4	0.19
EU-data EU-dis U.S.-work U.S.-scale	59.8	35.3	24.3	15.0	6.7	0.33
EU-data EU-dis U.S.-work EU-scale	61.8	39.3	27.8	17.7	8.2	0.33
EU-data EU-dis EU-work U.S.-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

9.8 shows simulated correlations between the binary outcomes employment (employment = 1 and nonemployment = 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 U.S. scales are used, because of justification bias, which is significant in the United States and absent in the European Union. 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 U.S. 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 European Union than in the United States.

9.8 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.

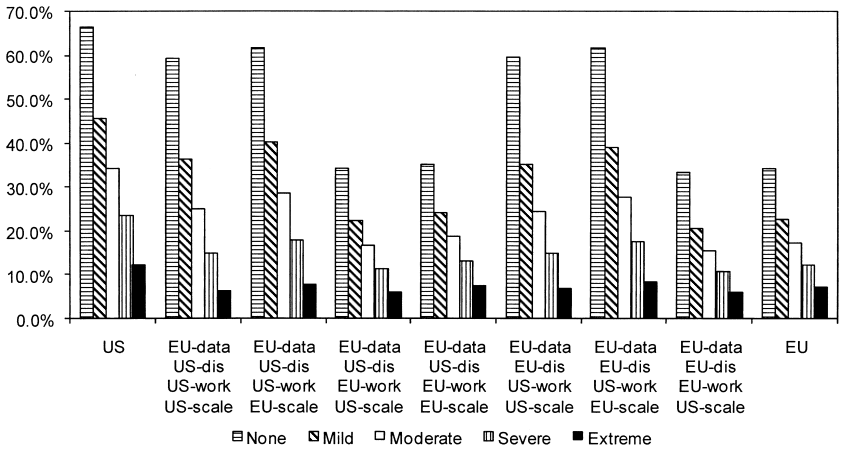

Fig. 9.5 Percent working by disability category

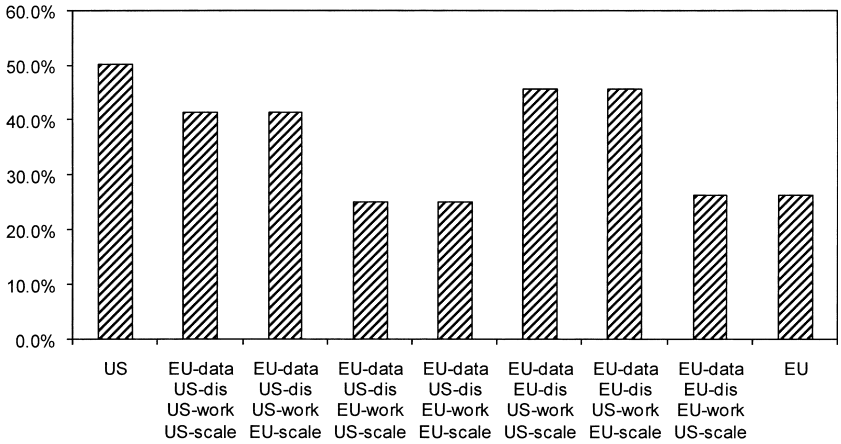

Fig. 9.6 Employment rates

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

Appendix A

Tables

Table 9A.1 **Responses to pain work disability vignettes**

	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	U.S.
Pain 1										
None	1.96	0.65	0.41	4.03	2.12	1.51	1.56	1.28	2.05	1.53
Mild	8.91	4.80	11.65	9.33	4.48	4.50	9.14	6.41	8.02	8.54
Moderate	35.08	22.56	26.00	30.27	28.55	13.98	40.68	31.64	31.25	34.99
Severe	47.77	62.20	41.12	43.47	36.49	53.16	43.83	42.00	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										
None	2.28	1.81	3.40	6.58	2.63	0.94	5.29	2.90	3.72	24.97
Mild	26.77	16.38	28.18	31.47	56.23	9.48	33.31	44.83	29.04	52.90
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	0.50	6.46	0.98	0.87	2.02	1.11
Total	100	100	100	100	100	100	100	100	100	100
Pain 3										
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	4.70	5.86	6.62	3.72
Moderate	22.06	25.41	23.77	33.73	35.78	37.14	25.70	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	9.50	20.41	16.91	22.53	7.69	6.68	12.05	12.90	23.62
Total	100	100	100	100	100	100	100	100	100	100

Table 9A.2 **Responses to affect disability vignettes**

	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	U.S.
Affect 1										
None	3.78	1.12	3.78	6.27	3.94	5.32	5.28	5.54	4.30	22.12
Mild	20.07	22.62	16.77	32.91	25.37	28.69	21.59	29.44	24.40	31.15
Moderate	48.64	45.91	43.70	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	21.70	22.56	22.12	13.26
Extreme	2.71	1.62	10.07	2.25	3.89	1.03	1.28	0.91	2.50	1.26
Total	100	100	100	100	100	100	100	100	100	100
Affect 2										
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	44.40	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	3.40	6.04	2.03	5.13	7.80	1.94	3.52	2.60	1.46
Total	100	100	100	100	100	100	100	100	100	100
Affect 3										
None	8.58	3.41	10.81	8.87	5.90	0.91	13.09	10.87	8.33	39.54
Mild	46.20	35.90	33.06	48.36	68.90	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	1.70	0	10.45	0.74	0.22	1.16	0.87
Total	100	100	100	100	100	100	100	100	100	100

Table 9A.3 **Responses to heart disease work disability vignettes**

	Germany	Spain	Greece	Italy	Netherlands	Sweden	France	Belgium	Total EU	U.S.
CVD1										
None	3.10	0.63	1.41	4.08	2.47	0.15	4.11	3.30	2.88	7.84
Mild	12.78	5.10	13.44	19.61	31.10	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	37.60	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										
None	3.31	0.30	0.60	4.47	3.76	1.95	3.62	2.46	2.98	8.16
Mild	11.27	4.09	5.16	20.73	26.70	24.30	9.95	18.89	13.47	20.46
Moderate	43.46	29.74	17.20	38.65	42.90	47.68	39.50	45.51	38.50	40.88
Severe	39.15	54.30	49.30	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	2.70	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										
None	3.99	0.45	1.86	5.31	3.04	0.15	2.14	3.83	3.12	3.10
Mild	3.20	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	11.40	17.34	29.66	19.80	24.99
Severe	53.39	57.53	33.63	36.62	32.80	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 9A.4 Parameter estimates; model with DIF

		Parameter	Standard error	*t*-value
		Disability equation		
U.S.				
	constant	–0.457*	0.131	3.490
	Female	–0.020	0.049	0.416
	Married/LT	–0.136*	0.052	2.645
	Educyrs	–0.043*	0.008	5.563
	Heart prob	0.437*	0.054	8.129
	Lung dis	0.391*	0.077	5.072
	High blood	0.114*	0.049	2.355
	Diabetes	0.241*	0.061	3.942
	Pain	0.384*	0.050	7.635
	Arthritis	0.342*	0.049	6.969
	Cancer	0.185*	0.069	2.674
	Cesd score	0.139*	0.013	11.147
	Obese	0.130*	0.053	2.434
	Age 58–64	0.141*	0.065	2.177
	Age 65–71	0.112+	0.063	1.766
	Age 72+	0.372*	0.072	5.157
Interaction EU				
	Female*EU	–0.090	0.062	1.451
	Marr/LT*EU	0.084	0.068	1.245
	Educyrs*EU	0.036*	0.008	4.574
	Heart*EU	0.013	0.075	0.172
	Lung*EU	–0.044	0.097	0.449
	Highbl*EU	–0.047	0.063	0.744
	Diabetes*EU	–0.011	0.087	0.122
	Pain*EU	0.067	0.064	1.052
	Arthr*EU	0.008	0.069	0.120
	Cancer*EU	0.235*	0.102	2.315
	Cesd*EU	0.148*	0.021	7.137
	Obese*EU	0.035	0.072	0.481
	Age 58–64*EU	–0.002	0.082	0.026
	Age 65–71*EU	0.170*	0.083	2.054
	Age 72+*EU	0.116	0.090	1.289
Country dummies				
	Germany	–0.913*	0.153	5.949
	Sweden	–1.130*	0.152	7.441
	Netherlands	–0.993*	0.151	6.573
	Spain	–1.279*	0.149	8.608
	Italy	–1.112*	0.152	7.292
	France	–1.244*	0.145	8.560
	Greece	–1.767*	0.149	11.875
	Belgium	–0.891*	0.150	5.955
		Threshold 1		
U.S.				
	const thrh	0	0	0
	Female	0.070*	0.016	4.315
	Married/LT	–0.011	0.017	0.684

(*continued*)

Table 9A.4 (continued)

		Parameter	Standard error	*t*-value
	Educyrs	–0.012*	0.003	4.651
	Heart prob	–0.023	0.020	1.174
	Lung dis	–0.004	0.030	0.136
	High blood	–0.014	0.017	0.863
	Diabetes	–0.068*	0.019	3.585
	Pain	–0.053*	0.018	2.889
	Arthritis	–0.017	0.017	0.995
	Cancer	0.044+	0.025	1.763
	Cesd score	–0.013*	0.005	2.751
	Obese	–0.008	0.017	0.489
	Age 58–64	–0.021	0.024	0.883
	Age 65–71	–0.037+	0.021	1.797
	Age 72+	–0.108*	0.024	4.418
Interaction EU				
	Female*EU	–0.077*	0.022	3.514
	Marr/LT*EU	0.021	0.024	0.907
	Educyrs*EU	0.007*	0.003	2.502
	Heart*EU	0.012	0.030	0.404
	Lung*EU	0.052	0.039	1.343
	Highbl*EU	–0.013	0.023	0.559
	Diabetes*EU	0.156*	0.029	5.376
	Pain*EU	0.014	0.025	0.577
	Arthr*EU	0.023	0.027	0.838
	Cancer*EU	–0.085*	0.038	2.254
	Cesd*EU	–0.005	0.009	0.508
	Obese*EU	0.061*	0.027	2.273
	Age 58–64*EU	0.033	0.031	1.080
	Age 65–71*EU	0.061*	0.029	2.126
	Age 72+*EU	0.093*	0.033	2.856
Country dummies				
	Germany	–0.639*	0.054	11.797
	Sweden	–0.657*	0.062	10.586
	Netherlands	–0.659*	0.051	12.993
	Spain	–0.915*	0.059	15.412
	Italy	–0.527*	0.051	10.332
	France	–0.575*	0.051	11.304
	Greece	–0.662*	0.057	11.715
	Belgium	–0.630*	0.054	11.695
	log(threshold 2 – threshold 1)			
U.S.				
	const thrh	–0.726*	0.048	15.162
	Female	0.040*	0.018	2.206
	Married/LT	0.071*	0.020	3.629
	Educyrs	0.017*	0.003	6.313
	Heart prob	0.017	0.022	0.751
	Lung dis	–0.030	0.032	0.928
	High blood	–0.008	0.018	0.464
	Diabetes	–0.002	0.024	0.101

Table 9A.4 (continued)

		Parameter	Standard error	*t*-value
	Pain	0.020	0.019	1.091
	Arthritis	0.021	0.019	1.122
	Cancer	0.014	0.026	0.547
	Cesd score	–0.008	0.005	1.586
	Obese	–0.021	0.020	1.078
	Age 58–64	0.016	0.025	0.640
	Age 65–71	–0.010	0.024	0.402
	Age 72+	0.048+	0.028	1.680
Interaction EU				
	Female*EU	0.009	0.025	0.369
	Marr/LT*EU	–0.061*	0.028	2.184
	Educyrs*EU	–0.013*	0.003	4.333
	Heart*EU	–0.041	0.035	1.184
	Lung*EU	0.045	0.044	1.036
	Highbl*EU	0.025	0.026	0.944
	Diabetes*EU	–0.098*	0.038	2.549
	Pain*EU	0.013	0.026	0.486
	Arthr*EU	–0.023	0.029	0.775
	Cancer*EU	0.043	0.043	1.005
	Cesd*EU	–0.001	0.010	0.072
	Obese*EU	–0.055+	0.030	1.796
	Age 58–64*EU	0.020	0.033	0.596
	Age 65–71*EU	0.048	0.034	1.440
	Age 72+*EU	–0.012	0.038	0.329
Country dummies				
	Germany	0.338*	0.062	5.435
	Sweden	0.180*	0.069	2.598
	Netherlands	0.665*	0.058	11.478
	Spain	0.447*	0.063	7.140
	Italy	0.441*	0.058	7.539
	France	0.289*	0.058	5.012
	Greece	0.238*	0.060	3.943
	Belgium	0.523*	0.059	8.926
	log(threshold 3 – threshold 2)			
U.S.				
	const thrh	–0.836*	0.049	17.019
	Female	0.012	0.019	0.627
	Married/LT	0.062*	0.020	3.028
	Educyrs	0.019*	0.003	6.497
	Heart prob	–0.008	0.022	0.380
	Lung dis	0.026	0.031	0.840
	High blood	–0.003	0.019	0.149
	Diabetes	0.027	0.024	1.137
	Pain	0.036+	0.019	1.916
	Arthritis	0	0.020	0.012
	Cancer	0.014	0.028	0.505
	Cesd score	–0.014*	0.005	2.800
	Obese	–0.022	0.021	1.068

(*continued*)

Table 9A.4 (continued)

		Parameter	Standard error	*t*-value
	Age 58–64	0.019	0.027	0.717
	Age 65–71	0.039	0.025	1.559
	Age 72+	0.126*	0.028	4.489
Interaction EU				
	Female*EU	–0.039+	0.022	1.742
	Marr/LT*EU	–0.060*	0.025	2.390
	Educyrs*EU	–0.018*	0.003	6.112
	Heart*EU	0.005	0.029	0.188
	Lung*EU	–0.047	0.037	1.281
	Highbl*EU	0.046*	0.023	1.965
	Diabetes*EU	–0.065*	0.032	2.069
	Pain*EU	0.023	0.023	1.010
	Arthr*EU	–0.039	0.025	1.543
	Cancer*EU	0.007	0.037	0.175
	Cesd*EU	0.026*	0.007	3.526
	Obese*EU	–0.014	0.027	0.528
	Age 58–64*E	0.012	0.031	0.379
	Age 65–71*E	–0.023	0.030	0.779
	Age 72+*EU	–0.066*	0.033	1.992
Country dummies				
	Germany	0.586*	0.055	10.590
	Sweden	0.285*	0.057	5.031
	Netherlands	0.478*	0.055	8.708
	Spain	0.550*	0.055	10.085
	Italy	0.558*	0.054	10.336
	France	0.589*	0.053	11.021
	Greece	0.395*	0.054	7.268
	Belgium	0.521*	0.054	9.577
	log(threshold 4 – threshold 3)			
U.S.				
	const thrh	–0.584*	0.052	11.278
	Female	–0.072*	0.020	3.597
	Married/LT	0.043*	0.021	2.020
	Educyrs	0.024*	0.003	7.703
	Heart prob	0.004	0.023	0.185
	Lung dis	0.001	0.035	0.030
	High blood	–0.011	0.020	0.554
	Diabetes	–0.055*	0.026	2.142
	Pain	0.002	0.022	0.104
	Arthritis	0.030	0.021	1.397
	Cancer	0.051+	0.031	1.661
	Cesd score	–0.016*	0.005	3.038
	Obese	–0.018	0.022	0.799
	Age 58–64	0.050+	0.028	1.742
	Age 65–71	0.051+	0.027	1.899
	Age 72+	0.080*	0.030	2.704
Interaction EU				
	Female*EU	0.047+	0.024	1.928
	Marr/LT*EU	–0.042	0.026	1.583

Table 9A.4 (continued)

		Parameter	Standard error	*t*-value
	Educyrs*EU	–0.022*	0.003	7.002
	Heart*EU	–0.009	0.031	0.293
	Lung*EU	–0.017	0.043	0.385
	Highbl*EU	–0.012	0.025	0.482
	Diabetes*EU	0.047	0.034	1.377
	Pain*EU	–0.021	0.026	0.806
	Arthr*EU	–0.026	0.028	0.929
	Cancer*EU	–0.004	0.043	0.098
	Cesd*EU	0.027*	0.008	3.398
	Obese*EU	–0.058*	0.028	2.039
	Age 58–64*EU	–0.001	0.034	0.023
	Age 65–71*EU	–0.016	0.033	0.492
	Age 72+*EU	0.017	0.035	0.472
Country dummies				
	Germany	0.560*	0.059	9.510
	Sweden	0.419*	0.058	7.171
	Netherlands	0.134*	0.059	2.261
	Spain	0.637*	0.056	11.288
	Italy	0.160*	0.059	2.683
	France	0.551*	0.057	9.741
	Greece	0.217*	0.057	3.821
	Belgium	0.313*	0.058	5.381
	Vignette dummies			
	d vig pain	1.003*	0.044	22.666
	d vig pain	0.202*	0.042	4.769
	d vig pain	1.145*	0.045	25.561
	d vig aff	0.372*	0.042	8.780
	d vig aff	0.424*	0.043	9.926
	d vig aff	0.019	0.042	0.454
	d vig cvd	0.641*	0.043	14.904
	d vig cvd	0.720*	0.043	16.620
	d vig cvd	1.167*	0.045	25.986
	Standard deviations of errors / unobserved heterogeneity			
	sig selfr	1	(Normalization)	
	sig thres	0.423*	0.007	63.621
	sig vignet	0.513*	0.007	69.770

Note: * = significant at two-sided 5 percent level; + = significant at two-sided 10 percent level.

Table 9A.5 Parameter estimates; model without DIF

		Parameter	Standard error	*t*-value
		Disability equation		
U.S.				
	constant	–0.135	0.107	1.261
	Female	–0.103*	0.043	2.377
	Married/LT	–0.179*	0.045	4.003
	Educyrs	–0.046*	0.007	6.957
	Heart prob	0.487*	0.048	10.232
	Lung dis	0.421*	0.068	6.175
	High blood	0.145*	0.044	3.312
	Diabetes	0.325*	0.052	6.238
	Pain	0.431*	0.045	9.625
	Arthritis	0.352*	0.044	8.073
	Cancer	0.142*	0.061	2.321
	Cesd score	0.171*	0.011	15.952
	Obese	0.163*	0.047	3.464
	Age 58–64	0.154*	0.058	2.636
	Age 65–71	0.130*	0.056	2.315
	Age 72+	0.461*	0.064	7.194
Interaction EU				
	Female*EU	–0.014	0.058	0.245
	Marr/LT*EU	0.113+	0.063	1.810
	Educyrs*EU	0.041*	0.007	6.052
	Heart*EU	–0.032	0.070	0.453
	Lung*EU	–0.139	0.090	1.543
	Highbl*EU	–0.068	0.059	1.152
	Diabetes*EU	–0.115	0.079	1.456
	Pain*EU	0.005	0.060	0.076
	Arthr*EU	–0.019	0.065	0.292
	Cancer*EU	0.260*	0.098	2.643
	Cesd*EU	0.114*	0.020	5.815
	Obese*EU	0.008	0.067	0.114
	Age 58–64*EU	–0.015	0.077	0.197
	Age 65–71*E	0.132+	0.077	1.709
	Age 72+*EU	0.012	0.083	0.150
Country dummies				
	Germany	–0.497*	0.139	3.571
	Sweden	–0.600*	0.135	4.433
	Netherlands	–0.609*	0.137	4.445
	Spain	–0.642*	0.134	4.791
	Italy	–0.790*	0.137	5.768
	France	–0.840*	0.130	6.449
	Greece	–1.254*	0.133	9.449
	Belgium	–0.535*	0.135	3.969
Threshold parameters				
	const thrh	0	0	0
	const thrh	–0.299*	0.013	22.472
	const thrh	–0.290*	0.013	21.482
	const thrh	–0.109*	0.013	8.164

Table 9A.5 (continued)

	Parameter	Standard error	*t*-value
Vignette dummies			
d vig pain	1.621*	0.024	68.127
d vig pain	0.747*	0.015	48.776
d vig pain	1.779*	0.025	70.588
d vig aff	0.928*	0.017	55.996
d vig aff	0.982*	0.018	55.413
d vig aff	0.563*	0.014	39.134
d vig cvd	1.217*	0.019	63.722
d vig cvd	1.309*	0.020	65.060
d vig cvd	1.798*	0.025	70.885
Standard deviations of error terms			
sig selfr	1	(normalization)	
sig vignet	0.728*	0.009	77.316
log-likelihood	–96584.3457		

Note: * = significant at two-sided 5 percent level; + = significant at two-sided 10 percent level.

Table 9A.6 **Parameter estimates for work disability equation**

	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

(*continued*)

Table 9A.6 (continued)

	Model without DIF		Model with DIF	
	Coeff.	St. error	Coeff.	St. error
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
Age 58–64*EU	–0.018	0.077	0.004	0.079
Age 65–71*EU	0.122	0.077	0.161*	0.080
Age 72+*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
Netherlands	–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*	0.130	–1.023*	0.136
Greece	–1.260*	0.133	–1.690*	0.139
Belgium	–0.542*	0.135	–0.654*	0.141

Note: * = significant at two-sided 5 percent level; + = significant at two-sided 10 percent level.

Table 9A.7 **Thresholds equation (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	Age 58–64*EU	0.034	0.028
Age 65–71	0.020	0.016	Age 65–71*EU	0.039	0.027
Age 72+	0.021	0.020	Age 72+*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	Netherlands	–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

Note: * = significant at two-sided 5 percent level; + = significant at two-sided 10 percent level.

Table 9A.8 **Vignette equation (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

Note: * = significant at two-sided 5 percent level; + = significant at two-sided 10 percent level.

Table 9A.9 **Employment equation**

	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

(*continued*)

Table 9A.9 (continued)

	Model without DIF		Model with DIF	
	Coeff.	St. error	Coeff.	St. error
Obese*EU	–0.296*	0.097	–0.284*	0.094
Age 58–64*EU	–0.399*	0.098	–0.406*	0.097
Age 65–71*E	–1.288*	0.133	–1.296*	0.130
Age 72+*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
Netherlands	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		–90792.80	

Note: * = significant at two-sided 5 percent level; + = significant at two-sided 10 percent level.

Table 9A.10 **Employment equation without objective health conditions**

	Model without DIF		Model with DIF	
	Coeff.	St. error	Coeff.	St. error
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
Age 58–64*EU	–0.391*	0.094	–0.398*	0.092
Age 65–71*EU	–1.277*	0.125	–1.287*	0.125
Age 72+*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
Netherlands	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		–90815.50	

Note: * = significant at two-sided 5 percent level; + = significant at two-sided 10 percent level.

Appendix B

The SHARE and HRS Work Disability Vignettes and Self-Reports SHARE

Work-Limiting Disabilities

Three domains: affect, pain, cardiovascular 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 three 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, and so forth. Version 1 has the order of the questions as just given. 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:

1a. 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

The 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. The SHARE vignette CVD-2 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.

References

Böckerman, P., and P. Ilmakunnas. 2009. Unemployment and self-assessed health: Evidence from panel data. *Health Economics* 18 (2): 161–79.

Boersch-Supan, A. 2007. Work disability, health, and incentive effects. Mannheim Research Institute for the Economics of Aging (MEA), University of Mannheim. Working Paper.

Bound, J. 1991. Self-reported versus objective measures of health in retirement models. *Journal of Human Resources* 26 (1): 106–38.

Croda, E., and J. Skinner. 2009. Disability Insurance and health in Europe and the U.S. Paper presented at Netspar 2009 Annual Conference, The Hague, April. Available at: http://www.netspar.nl/events/2009/annual/paperskinner.pdf.

Eurostat. 2001. *Disability and social participation in Europe.* Luxembourg: Office for Official Publications of the European Communities.

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 U.S. and the Netherlands. *American Economic Review* 97 (1): 461–73.

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. In *Research findings in the economics of aging,* ed. D. Wise, 63–92. Chicago: 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): 567–83.

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): 432–41.

Organization for Economic Cooperation and Development. 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. Available at: http://www.rand.org/pubs/working_papers/WR501/.

Comment Angus Deaton

I organize my comments around two facts or sets of facts that are at the core of this chapter:

- Fact 1: In matching surveys in Europe and the United States, people are asked to rate their own disability into five categories. The distribution of reports over those categories is very similar in Europe and the United States.
- Fact 2: In matching surveys in Europe and the United States, people are asked to rate other people's disability (using vignettes) into five categories. The distribution of reports over those categories is quite different in the two places; in particular, people in the United States are "tougher" in that they are less willing to admit that some conditions are disabling.

Kapteyn, Smith, and van Soest (henceforward KSvS) argue that fact 2 shows that Americans and Europeans use different scales to rank all disabilities, whether others or their own, a behavior that is known in the literature as "differential item functioning," or DIF for short. Once DIF is recognized, and the vignettes used to reinterpret the original responses in fact 1, we come to the conclusion that, in fact, the distribution of disability is worse in the United States than in Europe. I call this interpretation 1:

- Interpretation 1: Disability is worse in the United States than in Europe, but Americans have tougher standards, and so report the same levels as Europeans.

Fact 1 is misleading because Americans are tougher on themselves than are Europeans. The chapter makes a further contribution by allowing the extent of DIF to depend on whether or not someone is out of work—people might be less tough on themselves if they are not working, in effect a "justification bias." Their estimates suggest that this effect is present in the United States, but not in Europe. They conclude that "Americans use health as a justification for not working, whereas Europeans do not feel the need to do so."

Angus Deaton is the Dwight D. Eisenhower Professor of International Affairs and professor of economics and international affairs at the Woodrow Wilson School of Public and International Affairs and the Economics Department at Princeton University, and a research associate of the National Bureau of Economic Research.