Volume Title: Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly

Volume Author/Editor: David M. Cutler and David A. Wise, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-13231-5

Volume URL: http://www.nber.org/books/cutl08-1

Conference Date: October 8-11, 2004

Publication Date: January 2009

Chapter Title: Work Disability is a Pain in the ****, Especially in England, the Netherlands, and the United States

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

Chapter URL: http://www.nber.org/chapters/c11117

Chapter pages in book: (251 - 293)

9

Work Disability is a Pain in the ****, Especially in England, the Netherlands, and the United States

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

9.1 Introduction

High and rising rates of work disability are a pervasive problem in many industrialized countries (Bound and Burkhauser 1999). But rates of reported work disability vary considerably across countries with similar levels of economic development and comparable medical technology and treatment. Institutional differences in eligibility rules or generosity of benefits no doubt contribute to explaining the differences in disability rolls Haveman, Halberstadt, and Burkhauser (1984). Recent survey data show that significant differences between countries are also found in self-reports of work limiting disabilities and in general health. In comparing such self-reports, account should be taken of measurement issues such as differences in question wordings, as well as differences between and within countries that may exist in the scales that are used in answering questions about work disability.

This chapter investigates in some depth one highly salient—and as it turns out quite important—reason for reporting work disability, which is the presence of some type of pain. Unlike many illnesses of middle age, pain prevalence is very high. It also varies considerably across such key

James Banks is deputy research director at the Institute for Fiscal Studies, and a professor of economics at University College London. Arie Kapteyn is a senior economist at RAND and program director of RAND Labor and Population. James P. Smith is a senior economist at RAND. Arthur van Soest is a senior economist at RAND and a professor of econometrics at Tilburg University.

This research was supported by National Institute on Aging grants P30 AG12810 (to the National Bureau of Economic Research [NBER]) and R01 AG19805 (to the RAND Corporation), and by a Mary Woodard Lasker Charitable Trust and Michael E. DeBakey Foundation grant to the NBER. We thank David Rumpel for his expert programming assistance.

demographic attributes as gender and education. Most importantly for this chapter, amongst all health conditions pain is the most important determinant of work disability.

A unique aspect of this research is that it has a distinct multinational component by using data from three countries: the United States, United Kingdom, and the Netherlands. These three countries differ in a couple of relevant dimensions—observed rates of self-reported work disability, and perhaps national norms about the appropriateness of not working when one is or one claims one is work disabled. However, the countries appear to have similar economic standards of living and similar levels of objectively measured health status of the population. For this reason, international comparisons may be particularly useful in understanding some of the most salient research issues that have dominated the scientific literature on work disability.

Data on pain and its relationship to work disability are not abundant in any of the three countries. In addition to relying on a diverse set of currently available health and economic surveys in each country that do contain relevant information on pain and work disability, we have also been able to remedy that deficiency with new data collection efforts. First, we have had access to some reasonably large Internet samples in two of our countries, allowing us to experiment along several dimensions. These samples are the CentERpanel for the Netherlands and the RAND HRS and RAND MS Internet panels for the United States. For example, we placed experimental disability modules (with alternative forms of disability questions, etc.) and a pain module into these panels. In addition, the recently fielded English Longitudinal Survey on Aging (ELSA) has a detailed set of questions on pain, work disability, and workplace accommodations.

Pain has a subjective as well as objective manifestation, as individuals with the same amount of pain may react to it in very different ways. Another aspect of this chapter is that we utilize the vignette methodology to evaluate—once again in an experimental setting—how people within the same country as well as across countries set thresholds that result in labeling some people work disabled while other people are not so described. Vignette questions have been applied successfully in recent work on international comparisons of health and work disability (King 2004; Kapteyn, Smith, and Van Soest 2007). In this chapter, we will use vignettes on pain to identify systematic differences in self-reported work disability in the Netherlands and the United States.

One reason why pain may have differential impacts on work disability in the three countries is that practices differ on how to limit the effects of pain on people's ability to function effectively in their lives, especially in the workplace. Two aspects of possible cross-country differences will be investigated—the use of medication to relieve pain and the availability of workplace accommodations that lessen its impact on the job.

The remainder of this paper is divided into five sections. Section 9.2 compares and evaluates the impact of some differences in wording of work disability questions both within and across countries on reports of work disability. Section 9.3 summarizes several salient differences and similarities in the type, severity, and duration of pain in our three countries. This section also documents the one-way and multivariate relationship between pain and self-reports of work disability in each country. Section 9.4 examines differences across countries in pain medication and workplace accommodations. Section 9.5 summarizes our results using the vignette methodology, and section 9.6 presents our conclusions.

9.2 Does the Form of the Question Matter?

It is an understatement that there is no agreed upon standard format for asking about work disability. Thus, it is not surprising that the format and wording of questions on work disability vary not only internationally but also across the major social science surveys within a country. For example, in the United States quite different questions are asked in the principal yearly government labor force survey (the Current Population Survey or CPS) and the principal yearly health survey (National Health Interview Survey or NHIS) (Burkhauser et al. 2002). To illustrate, the CPS question is:

Does anyone in the household have a health problem or disability which prevents them from working or which limits the kind or amount of work they can do? [If so,] who is that? (Anyone else?)

while the NHIS asks instead two questions:

Does any impairment or health problem now keep you from working at a job or business?

Are you limited in the kind of amount of work you can do because of any impairment?

To add to the potential domestic confusion, the work disability question in the HRS is

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

and for Panel Study of Income Dynamics (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?

In all cases, the answers permitted are *yes*, *no*, *don't know*, or *refuse* so that essentially a dichotomous disability scale can be created.

Some differences between the ways these questions are asked involve language. National Health Interview Survey and Health and Retirement

Study (HRS) use the term impairment; NHIS, HRS, and CPS use health problem; PSID contains only the phrase physical or nervous condition; while the word disability is only used explicitly in CPS. Another potentially important difference is that CPS first asks about anyone in the household and then in a follow-up inquires about whom that might be.

Not surprisingly, survey differences in the manner in which work disability questions are asked are not limited to the United States. For example, the basic work disability question in the Dutch CentERpanel is:

Do you have an impairment or health problem that limits you in the amount or kind of work you can do?

While this sounds very similar to the HRS question format, the possible answers are now arrayed on the following five point scale:

(1) no, not at all, (2) yes, I am somewhat limited, (3) yes, I am rather limited, (4) yes, I am severely limited, and (5) yes, I am very severely limited—I am not able to work.

Finally, in England the disability question used in the British Household Panel Survey (BHPS) is very similar but not identical to the HRS variant—"Does your health limit the type of work or the amount of work you can do?" While ELSA did not have a work disability question in wave 1, the designers placed the following question into the first follow-up: "Do you have any health problem or disability that limits the kind or amount of work you can do?"[1]

This wide variation in the form in which work disability questions are asked both within and between countries raises the question of how important this variation is in creating differences in reported rates of disability prevalence.

9.2.1 Reports of Disability Prevalence

In this project, we conducted several experiments to evaluate the impact of differences in question wording on reporting of disability prevalence. First, we placed the disability questions summarized above from the HRS, CPS, and NHIS into the RAND HRS Internet panel. This panel is based on a sample of about 2,700 respondents in the HRS 2002 wave who had Internet access and who expressed a willingness to participate in an experimental survey on the Internet. This panel allows us to test in a random experimental setting whether the alternative forms of these questions in these three prominent surveys lead to very different measures of disability prevalence using the same population of respondents. Moreover, the reasons for

1. If the answer to this question is yes, ELSA follows the HRS format by asking, "Is this a health problem or disability that you expect to last at least three months?"

Table 9.1 **Disability prevalence**

	Percentage of cases who report disability
NHIS	18.0
HRS	17.4
CPS	24.6
HRS nonmarried	23.5
CPS nonmarried	24.1
NHIS nonmarried	21.4

Note: Sample is from RAND HRS Internet sample.

any differences that emerge can be subsequently explored using the rich information available from the core HRS interviews.[2]

In the RAND HRS Internet panel, we conducted the following experiments: half of the sample was randomly assigned the NHIS form of the disability question while the other half received the CPS variant. To test for mode differences (the Internet versus the telephone in the prior wave), the full RAND HRS Internet sample received the normal HRS question. The principal results are contained in table 9.1.

Contrary to the speculation in the literature, there does not appear to be any difference in estimates of disability prevalence induced by the wordings of these alternative questions. The NHIS and HRS variants produce bang-on estimates. One complication in making these comparisons is that HRS staff has not coded the specific people affected in the CPS question. Fortunately, a fix is available by limiting the comparisons to nonmarried respondents. Table 9.1 shows that in this sample HRS, CPS, and NHIS produce remarkably similar sets of estimates about disability prevalence.

While the PSID disability question was not included in these experiments, one can compare PSID estimates of work disability prevalence with those obtained in the HRS for the same age group. In that case the PSID estimate of work disability was 28.7 percent while it was 26.8 percent in the HRS, about a 2 percentage point difference. This also does not seem to us to be a large difference, but this conclusion must be qualified by the fact that, unlike the numbers in table 9.1, this comparison is not a strict comparison of question wording only, as other factors such as sampling frames likely differ between the surveys in view of the fact that the HRS sample only includes respondents with Internet access.

Similarly, two other British surveys in addition to the British Household Survey (BHPS) ask work disability questions. For example, the Labor

2. The HRS respondents with Internet access are a selective sample of the population. However, since we are comparing within sample it seems unlikely that our results are very much affected by this selectivity.

Force Survey (LFS) first asks, "Do you have any health problems or disabilities that you expect will last for more than a year?" If the answer is yes, then respondents are asked in sequence, "Does this health problem affect the KIND of paid work that you might do?" and then, "or the AMOUNT of paid work that you might do?" The other survey is called the Family Resource Survey (FRS), which asks, "Some people are restricted in the amount or type of work they can do, because they have an injury, illness, or disability. Which of these statements comes closest to your own position at the moment? (1) Unable to work at the moment; (2) Restricted in amount or type of work I can do; (3) Not restricted in amount or type of work I can do." In spite of the difference in the manner in which these questions are asked, prevalence rates from the BHPS, LFS, and FRS are remarkably close.

Thus, in our view any conflicts that emerge amongst these surveys in estimates of the prevalence of the work disabled population appear not to be due to the form of the disability questions. One possible explanation is that the greater concentration on health content in the NHIS alerts their respondents to health issues and results in higher reporting of disability, although differences in sampling frames may be a more likely explanation.[3]

Our next set of experiments was conducted using the Dutch CentERpanel, which includes about 2,000 households who have agreed to respond to a set of questions every weekend over the Internet. Unlike the RAND HRS Internet panel, this Dutch sample is not restricted to households with their own Internet access. If they agree to participate and do not currently have Internet access, they are provided Internet access.[4] One advantage of the Dutch Internet panel is that these respondents had already answered many questions about their lives, including questions about their health, demographics, and labor force activity. In this project, we carried out a number of experiments over about a six-month period. These included the vignette experiments, which are reported on below, test-retest experiments, and experiments with question wording. The experiments took place mid-August, mid-October, and mid-December 2003.

For example, in the second round of the CentERpanel vignette disability experiments (mid-October 2003), we conducted another experiment about question wording. Randomly, half of CentERpanel respondents in the second wave of our vignette experiments were given the HRS disability

3. Some evidence is available from ELSA which experimented with placing the general health status questions before and after the detailed set of questions that inquired about a long list of possible health problems. There was some tendency to report better general health status when the questions were placed at the end but the principal difference was that there were fewer respondents at either tail of the five point general health scale when the questions were at the end.

4. Providing Internet access may require just a subscription with an Internet provider, but usually it involves the provision of a set-top box which is connected to a TV set and a telephone line to allow Internet access; if needed, a TV set is also provided.

question whereby one answered on a yes no basis to the disability question. In the first round (mid-August 2003), the same question had been asked with a five point response scale, as noted above. Given that the first and second waves of our experiments were only a few months apart so that disability reports should not change that much, for these respondents one can compare the answers to this question to that given on the five point scale a few months earlier.

The results are presented in table 9.2. For all but one row in the five point scale, the correspondence is remarkably close. Ninety-six percent of those who answered they were not at all disabled on the five point scale also said that they were not when using the HRS dichotomous scale. Similarly, more than 90 percent of Dutch respondents who said that they were more than somewhat limited replied that they had a work disability on the U.S. two point scale.

The ambiguity occurs within the somewhat limited category, which splits about fifty/fifty when offered an opportunity to simply respond yes or no about their work disability. These are people who are clearly on the margin in terms of their work disability problems. When offered a stark yes or no choice, some will resist disability labeling. But if given a more nuanced set of alternatives, they report some degree of disability.

Since this somewhat limited group represents just under a quarter of Dutch respondents, the implication is that reports of disability prevalence are considerably lower if the two-point scale is used in place of the five-point scale. Table 9.3 shows reported U.S. disability rates by age (from the PSID) alongside those in the U.K. (from the Labor Force Survey) and the Dutch disability rates using the five and two point scale obtained from CentERpanel. Especially during middle age, the Dutch have the highest rates of self-reported work disability, followed by the British, with the Americans having the lowest rates. While estimates of Dutch disability prevalence using the dichotomous scale are still much higher than that observed in the United States, a significant fraction of the disparity could be explained by the format of the disability scale. However, especially for middle age workers—say those between ages forty-five to sixty-four—

Table 9.2 **Correspondence between 5 and 2-point scale in Dutch panel**

5-point work limitations	Percentage in category	marginal Percentage disabled in 2-point scale
Not at all	61.8	4.3
Somewhat limited	22.5	56.1
Rather limited	9.9	91.2
Severely limited	2.2	93.1
Very severely limited	3.6	92.1

Source: Dutch CentERpanel.

Table 9.3 **Percentage with work disability by age—United States, United Kingdom, and the Netherlands**

	Age group				
	25–34	35–44	45–54	55–64	65+
United States	7.4	11.3	17.6	25.9	38.8
United Kingdom	9.1	12.4	19.4	30.8	n.a.
Netherlands					
5-point scale	25.7	30.3	42.7	44.2	53.6
U.S. 2-point scale	17.2	23.6	38.7	37.4	38.8

Notes: U.S. data are from PSID. U.K. data are from 2001 Labor Force Survey. Due to question routing, the fifty-five to sixty-four group contains women ages fifty-five to fifty-nine and men ages fifty-five to sixty-four. Netherlands data are from CentERpanel. Netherlands 5-point scale is based on report of any limitation. U.S. and U.K. use the 2-point scale. All data are weighted. n.a. = not available.

Dutch rates of reported work disability are still about 15 percentage points higher than those in the United States even when the same question is asked in both countries.

9.3 Pain and Work Disability

In this section, we discuss the central role played by pain as a potential determinant of work disability. The amount and type of pain information available differs in several ways across the countries we study. Rather than going straight to the lowest common denominator by restricting our analysis to information that is available and identical in all three countries, we take the alternative strategy of using the best information that each country has to offer. While comparability across countries will not be exact, this will still provide the most useful information about the relative importance of pain in affecting work disability.

More so than many specific diseases, pain has subjective and objective aspects. Objectively, in a reaction to a variety of stimuli, pain is started when energy is converted into electrical energy (nerve impulses) by sensory receptors called nociceptors. These neural signals are then transmitted to the spinal cord and brain, which perceives them as pain. Some pain medications or analgesics can inhibit nociception and thereby lessen or even eliminate the sensation of pain. Even without medication, individuals differ in how they access, interpret, and tolerate pain so that there may well be a significant subjective component to the reporting of pain, both within and across countries. As shown in the following paragraphs, pain also varies in its severity, duration, and location, all of which may have different implications for the tolerance and perception of pain and for work disability.[5]

5. See the web site of the American Pain Society. Accessed at http://www.ampainsoc.org

With this in mind, table 9.4 provides information about the prevalence and types of pain people experience in the United States, the Netherlands, and the United Kingdom, respectively. Unless otherwise indicated, all data in this table refer to individuals ages twenty-five and over. Pain prevalence rates are also stratified by gender, education, and age. Just like work disability, commonly used questions used to ascertain whether an individual has pain or not also vary a good deal in their format and wording, both across different surveys within countries and across countries. However, unlike the form of questions on work disability, the specific language used in pain questions appears to really matter a lot. For example, the most basic question asked in the National Health Interview Survey (NHIS) in the United States about pain was whether an individual had any recurring pain during the last twelve months, while the most comparable question in the Dutch CentERpanel was, "Are you often troubled by pain?" We will refer to this question form as the recurrent pain question.

Another common form in which pain questions are asked involves inquiring about the presence of pain in specific parts of the body from which an aggregate of pain can be deduced. The American and Dutch surveys used the same parts of the body—neck, back, face or jaw, joints, and headaches. The British survey only asks about migraines. However, these questions tend to ask about the presence of pain over shorter periods of time—for example, in the American NHIS the reference period used is the last three months, while in the Dutch panel the last thirty days are used. We will refer to this question form as the recent pain question.

The situation in England is more complicated. The 1999 British Household Survey (BHPS) contained the SF-36 questionnaire (Ware and Sherbourne 1992). As a consequence all respondents were asked, "How much bodily pain have you had during the past four weeks?" where the allowed responses follow a six-point scale: none, very mild, mild, moderate, severe, and very severe. In addition, a second item of the SF-36 (again delivered to all respondents) asks, "During the past four weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?" where five possible responses are allowed: not at all, a little bit, moderately, quite a bit, and extremely. This SF-36 questionnaire has not yet been delivered again to BHPS respondents. However, in the 2001 wave of the BHPS, respondents were asked, "Are you regularly troubled by pain?" a question that is quite similar to the one asked in the Dutch CentERpanel. Unfortunately, this question was only asked of respondents ages fifty and over. Those reporting yes to this question are asked how often they are troubled by pain (every day, at least once a week, once a month, less often), and how they would describe pain (mild, moderate, or severe). In summary, all BHPS respondents were asked in the 1999 wave a form of the recent pain question while BHPS respondents in the 2001 wave age fifty and over were asked a version of the recurrent pain question.

Table 9.4 **Prevalence of types of pain, ages 25+**

	All	Men	Women	Education low	Education med.	Education high	Ages 45+	Ages 45–64
			United States					
Recurring pain in last 12 months	19.6	17.3	21.5	23.2	20.8	15.1	23.7	23.9
Any pain in last 3 months	51.3	47.3	52.1	55.5	53.1	46.5	56.1	55.0
Neck pain	14.9	12.6	16.8	17.2	15.8	11.6	16.0	17.0
Jaw/face pain	4.7	2.8	6.4	5.3	5.1	3.6	4.6	5.2
Back pain	27.5	25.5	29.0	32.6	28.8	21.9	29.6	30.0
Joint pain	38.7	29.8	42.8	37.7	33.1	26.8	40.9	37.7
Severe headaches/migraines	14.9	9.2	19.9	16.9	15.8	11.7	12.4	15.2
			Netherlands					
Often troubled by pain	26.7	20.7	33.1	29.9		19.5	31.6	32.1
Any pain in last 30 days	58.9	51.8	66.4	60.5		55.5	60.5	60.2
Neck pain	20.6	16.2	25.3	22.1		17.3	21.7	23.9
Jaw/face pain	5.7	3.7	7.9	6.4		4.2	5.9	7.9
Back pain	32.9	28.9	37.1	35.9		26.1	34.1	32.6
Joint pain	37.4	34.1	40.8	40.4		30.5	44.3	42.3
Headaches/migraines	25.4	16.9	34.3	25.9		24.1	21.2	27.1
			United Kingdom					
Have mild pain or more in last 4 weeks	39.5	33.8	44.1	48.5	33.9	29.7	46.6	41.9
Have moderate pain or more in last 4 weeks	26.5	21.3	30.8	35.1	21.2	17.1	32.6	28.2
Migraines	8.8	4.7	12.2	8.8	9.7	7.1	7.9	9.6

Source: U.S.—National Health Interview Survey (NHIS) 2002. All places of pain are defined over the last three months except joint pain, which is defined over the last thirty days. Any pain in last three months includes the one-month joint pain. Netherlands—CentER panel, December 2004. Each of the specific types of pain are during the last thirty days and any pain in last thirty days means that you had at least one type. United Kingdom—British Household Panel Survey 1999. Units are in percents.

In all three countries, prevalence rates are considerably lower with the recurrent pain than in the recent pain formulation. For example, while one in five adult Americans report some form of recurring pain during the last year, about half of them report having pain somewhere during the last three months. Similarly, while a little more than a quarter of adult Dutch respondents said that they were often troubled by pain, sixty percent of them reported that they had some pain in some place during the last thirty days.

There are several possible reasons for this difference. First, the use of words such as recurring or often may imply a higher pain threshold, especially in its temporal duration, that recent pain questions cannot match. Reflecting a standard result from retrospective memory studies, recent pain may also be more likely to be recalled, thereby increasing its reported prevalence. Finally, any recent pain is calculated by going through specific types of pain like back pain, which because it is less vague and more specific, may stimulate recall. This is quite similar to findings that total consumption measures that are computed by asking about specific consumption items yield higher consumption totals than a catch-all single total consumption question (Browning, Crossley, and Weber 2003).

In whatever form the pain question is asked, there are several key similarities among the countries. In each country, women are much more likely to report that they suffer from pain than men are, pain prevalence declines significantly as education increases, and the age gradient in pain is actually quite muted. If we compare Dutch, Americans, and British using the more comparable recent pain formulation, prevalence levels of pain appear higher in the Netherlands than in the other two countries.

Table 9.4 also documents that pain in the joints and back pain are the most common types of pain that people report in both the Netherlands and the United States. All forms of pain, including joint and back pain, have very pronounced negative gradients across education groups. Finally, all types of pain are more prevalent among women than they are amongst men, and in all three countries, severe headaches or migraines appear to especially be a problem for women. For example, more than a third of Dutch women report that they suffer from headaches compared to less than one in six Dutch men.

Individuals also differ in the severity of the pain that they experience. Table 9.5 summarizes the respondents' assessments of the severity of the pain that they experience, with that assessment placed into three categories—light, moderate, and heavy. While the specific scales used to place individuals within these three groups differ between the countries, the patterns that emerge across groups are quite similar. In each country, there is a great deal of variation amongst people in how they evaluate the severity of the pain that they experience. Women are more likely to say that they experience more severe pain, and in all three countries less-educated individuals are more likely to state that their pain was not light.

Table 9.5 **Severity of joint pain in the United States, the Netherlands, and the United Kingdom, ages 25+ (percent distributions)**

	All	Men	Women	Education low	Education med.	Education high
			United States			
Light	27.6	31.7	24.2	17.1	25.1	42.1
Moderate	53.2	45.4	52.2	51.4	54.7	50.2
Heavy	19.3	14.0	23.5	31.5	20.3	7.6
			Netherlands			
Light	36.3	38.4	34.2	22.5	n.a.	30.6
Moderate	46.7	49.1	43.4	50.5	n.a.	42.1
Heavy	17.6	12.5	28.3	27.0	n.a.	27.2
			United Kingdom			
Light	52.7	58.0	49.0	44.6	58.6	64.8
Moderate	28.9	26.8	30.3	31.3	27.0	25.3
Heavy	18.4	15.2	20.7	24.1	14.4	9.9

Source: U.S.—National Health Interview Survey (NHIS) 2002. U.S. respondents were asked to rank their pain on a scale of 0–10 with 0 being *no pain* and 10 *very bad pain*. This numerical scale was converted as follows: 0–3 = *light*, 4–7 = *moderate*, 8–10 = *heavy*. Netherlands—CentERpanel, December 2004. Dutch respondents were asked to rank their pain into one of the three categories listed in this table. UK—1999 British Household Panel Survey.

Notes: Respondents were asked to rank from 0 to 5, where 0 = *no pain* in the last 4 weeks. Sample is those who do not report *no pain*. We convert that ranking as follows: 2–3 = *light*, 4 = *moderate*, 5–6 = *heavy*. UK respondents were asked to rank from 1 to 5. We convert that ranking as follows: 1–2 = *light*, 3 = *moderate*, 4–5 = *Heavy*. n.a. = not available.

Using the alternative forms of the definitions of pain used in tables 9.4 and 9.5, table 9.6 documents the relationship between the presence of pain and the report of a work disability. These simple cross-tabular relationships suggest that pain is a very powerful correlate of work disability. No matter which specific definition of pain is used, those who claim that they suffer from pain are much more likely to also say that they have a work disability. To illustrate using the recurrent pain question, Dutch respondents who say that they are often troubled with pain are almost four times as likely to say they are work disabled than those who do not have pain (64.9 percent compared to 16.9 percent). That difference is even larger among Americans (35.7 percent compared to 7.5 percent). Just as in the other two countries, work disability in the United Kingdom is around four times higher for those with general pain than for those without. And as in the other countries, when looking at specific pain, in this case migraines, the differences between those with and without such pain are still apparent although the relative risk of work disability is somewhat lower.

All forms of pain that we measured appear to be strongly associated with work disability. Recurrent pain appears to be somewhat more strongly associated with work disability, and among the alternative types of pain that

Table 9.6 **Work disability by presence of pain, ages 25+ (percent distributions)**

	All with pain	All without pain	Education low with pain	Education low without pain	Education med with pain	Education med without pain	Education high with pain	Education high without pain
				United States				
Recurring pain in last 12 months	35.7	7.5	52.4	17.0	35.8	7.3	21.4	2.9
Any pain in last 3 months	21.2	7.8	36.5	17.0	21.4	7.7	10.2	2.3
Neck pain	27.4	10.5	45.0	21.2	27.1	10.6	14.6	4.5
Jaw/face pain	31.7	12.1	52.2	22.7	32.6	12.2	12.5	5.5
Back pain	24.3	8.7	39.6	18.3	24.2	8.8	11.7	4.0
Joint pain	25.3	7.2	41.9	15.2	25.1	7.4	13.2	3.0
Severe headaches/migraines	22.7	11.3	40.1	22.2	22.5	11.5	9.6	5.2
Pain light	11.6	n.a.	26.1	n.a.	11.1	n.a.	7.9	n.a.
Pain moderate	24.8	n.a.	37.9	n.a.	25.0	n.a.	15.0	n.a.
Pain heavy	44.4	n.a.	55.2	n.a.	41.8	n.a.	29.3	n.a.
				Netherlands				
Often troubled by pain	64.9	16.9	66.9	18.0			58.0	14.7
Any pain in last 30 days	42.1	11.9	45.7	12.6			33.4	10.4
Neck pain	54.3	23.3	57.7	25.5			44.3	18.7
Jaw/face pain	66.3	27.5	70.1	30.1			53.1	21.8
Back pain	49.9	19.8	53.3	21.1			39.3	17.4
Joint pain	55.0	14.6	58.6	15.0			44.2	13.9
Headaches/migraines	42.3	25.4	46.1	27.9			33.0	20.0
Pain light[a]	27.0	n.a.	28.7	n.a.			23.3	n.a.
	16.1		14.3				19.2	

(*continued*)

Table 9.6 (continued)

	All with pain	All without pain	Education low with pain	Education low without pain	Education med with pain	Education med without pain	Education high with pain	Education high without pain
Pain moderate[a]	65.3	n.a.	68.0	n.a.			54.8	n.a.
	39.2		42.5				30.7	
Pain heavy[a]	85.8	n.a.	89.2	n.a.			75.5	n.a.
	66.3		69.3				55.9	
			United Kingdom					
Have mild pain or more in last 4 weeks	40.7	9.7	50.6	14.9	31.1	6.4	25.5	7.2
Have moderate pain or more in last 4 weeks	49.5	12.0	57.7	18.4	39.9	8.0	34.2	8.2
Severe headaches/migraines	30.1	21.1	40.7	31.4	22.7	13.9	19.8	12.1
Pain light	9.7	n.a.	14.9	n.a.	6.4	n.a.	7.2	n.a.
Pain moderate	22.8	n.a.	31.8	n.a.	16.5	n.a.	13.7	n.a.
Pain heavy	47.8	n.a.	55.8	n.a.	38.8	n.a.	34.7	n.a.

Source: U.S.—National Health Interview Survey (NHIS) 2002. All places of pain are defined over the last three months except joint pain which is defined over the last thirty days. Any pain in last three months includes the one-month joint pain. Each cell presents the percentage of respondents with work disability. For instance, the entry 35.7 indicates that among those with recurring pain in the last twelve months, 35.7 percent reports to be work disabled; the entry 7.5 indicates that among those who do not report a recurring pain in the last twelve months, only 7.5 percent reports to be work disabled. Netherlands—CentERpanel, December 2004. Each of the specific types of pain are during the last thirty days and any pain in last thirty days means at least one type. United Kingdom—British Household Panel Survey 1999.

Notes: n.a. = not available.

[a]First number: pain in joints only; second number: most serious types of pain (of the five types: neck pain, jaw/face pain, back pain, joint pain, headaches/migraines).

are included in our surveys, joint pain appears to have the strongest association. Not surprisingly, respondents' report of the severity of pain is quite crucial for whether a work disability is also reported. For example, among Americans those with heavy pain are four times more likely to say that they are work disabled than those who categorize their pain as only light. If anything, this difference is even larger in the Dutch sample. Even after one controls for the degree of pain severity, those in lower education groups are much more likely to report that the pain results in a work disability.

Pain is certainly not the only thing that matters for work disability. Therefore, we next estimated probit models of the probability that a respondent reported having a work disability. The American, Dutch, and British models are presented in table 9.7. In addition to variables that capture some aspect of pain, these models include measures of a standard set of demographic attributes (gender, education, marital status, and age) as well as a list of as many chronic health conditions that are available in the data (hypertension, diabetes, cancer, diseases of the lung, heart problems, stroke, emotional problems, and arthritis).

In each country, three variants of the model were estimated—one with an indicator of pain, the second which categorizes the severity of this pain, and the third of which includes indicators of the location of pain. As mentioned above, places of pain are not available in the United Kingdom, so in its stead we include a second variant where the pain threshold is moderate pain or worse. All tables list estimated coefficients, derivatives, and z values of estimated differences from zero in the three countries.

We first discuss the nonpain variables in these models. The Dutch samples are much smaller than those available in the other two countries. Putting that caveat aside and given the differences in the institutional context in each country, and especially the diverse manner in which the pain questions are formulated, one is struck by the basic similarity in model estimates across the three countries. In these models in all three countries, work disability falls significantly with education level, rises with age, and is lower among married respondents. The only demographic difference that emerges concerns gender. In the United States work disability is lower among women (statistically significant), while it is not different by gender in the other two countries. Finally, all the health problems included in these models appear generally to have independent and statistically significant effects on work disability.

Pain turns out to be the most important predictor of work disability in all three countries. Moreover, pain—in each of the forms in which we measure it (place of pain and its severity)—is a statistically significant independent predictor of work disability.

Our goal with these models is twofold—to uncover the principal factors that led to a report of work disability and to isolate the sources of the international difference in reported work disability. To see how we accomplish

Table 9.7 **Probits for work disability**

	Coefficient	DF/dX	Coefficient	DF/dX	Coefficient	DF/dX
			United States			
High blood pressure	0.149	0.025	0.137	0.024	0.131	0.022
	(6.06)**	(6.06)**	(5.55)**	(5.55)**	(5.28)**	(5.28)**
Diabetes	0.323	0.063	0.308	0.060	0.317	0.061
	(9.23)**	(9.23)**	(8.79)**	(8.79)**	(9.01)**	(9.01)**
Cancer	0.238	0.044	0.240	0.045	0.221	0.040
	(6.71)**	(6.71)**	(6.75)**	(6.75)**	(6.18)**	(6.18)**
Lung disease	0.390	0.079	0.391	0.080	0.347	0.068
	(10.26)**	(10.26)**	(10.20)**	(10.20)**	(8.98)**	(8.98)**
Heart problems	0.391	0.077	0.403	0.081	0.380	0.074
	(13.25)**	(13.25)**	(13.60)**	(13.60)**	(12.77)**	(12.77)**
Stroke	0.585	0.133	0.596	0.138	0.584	0.131
	(10.46)**	(10.46)**	(10.56)**	(10.56)**	(10.32)**	(10.32)**
Arthritis	0.465	0.049	0.368	0.069	0.317	0.057
	(18.85)**	(18.85)**	(13.73)**	(13.73)**	(11.81)**	(11.81)**
Emotional problems	0.694	0.159	0.692	0.160	0.629	0.138
	(22.78)**	(22.78)**	(22.53)**	(22.53)**	(19.95)**	(19.95)**
Pain	0.410	0.072				
	(17.93)**	(17.93)**				
Pain light			0.038	0.006		
			(0.94)	(0.94)		
Pain moderate			0.369	0.072		
			(12.64)**	(12.65)**		
Pain heavy			0.704	0.167		
			(17.93)**	(17.93)**		
Neck pain					0.164	0.028
					(5.33)**	(5.33)*
Back pain					0.289	0.051
					(11.40)**	(11.40)**
Jaw pain					0.156	0.027
					(3.37)**	(3.37)**
Headache					0.171	0.030
					(5.49)**	(5.49)**
Joint pain					0.292	0.050
					(11.38)**	(11.38)**
Female	–0.136	–0.025	–0.130	–0.022	–0.150	–0.024
	(6.07)**	(6.07)**	(5.79)**	(5.79)**	(6.55)**	(6.55)**
Education med.	–0.237	–0.040	–0.232	–0.039	–0.238	–0.039
	(8.88)**	(8.88)**	(8.30)**	(8.30)**	(8.34)**	(8.34)**
Education high	–0.538	–0.074	–0.511	–0.071	–0.529	–0.071
	(14.50)**	(14.50)**	(13.79)**	(13.79)**	(14.06)**	(14.06)**
Age 35–44	0.271	0.049	0.249	0.045	0.260	0.046
	(6.72)**	(6.72)**	(6.19)**	(6.19)**	(6.36)**	(6.36)**
Age 45–54	0.445	0.087	0.401	0.078	0.430	0.082
	(11.11)**	(11.11)**	(10.02)**	(10.02)**	(10.58)**	(10.58)**
Age 55–64	0.606	0.130	0.548	0.116	0.604	0.127
	(14.17)**	(14.17)**	(12.85)**	(12.85)**	(13.95)**	(13.95)**
Age 65+	0.526	0.010	0.445	0.087	0.549	0.108
	(12.23)**	(12.23)**	(10.42)**	(10.42)**	(12.53)**	(12.53)**

Table 9.7 (continued)

	Coefficient	DF/dX	Coefficient	DF/dX	Coefficient	DF/dX
Married	–0.412	–0.068	–0.408	–0.068	–0.412	–0.067
	(18.14)**	(18.14)**	(17.99)**	(17.99)**	(18.03)**	(18.03)**
Constant	–1.633		–1.526		–1.658	
	(34.00)**		(32.50)		(34.40)**	
Observations	27,684		27,684		27,684	
Observed *p*	0.146		0.146		0.146	
Log likelihood	–8,541.1		–8,494.0		–8.403.3	
			Netherlands			
High blood pressure	0.007	0.002	–0.028	0.008	0.011	0.003
	(0.07)	(0.07)	(0.28)	(0.28)	(0.11)	(0.11)
Diabetes	0.531	0.180	0.514	0.173	0.602	0.205
	(2.85)**	(2.85)**	(2.70)**	(2.70)**	(3.25)**	(3.25)**
Cancer	0.260	0.082	0.127	0.038	0.265	0.082
	(1.31)	(1.31)	(0.62)	(0.62)	(1.32)	(1.32)
Lung disease	0.467	0.156	0.513	0.172	0.433	0.141
	(2.79)**	(2.79)**	(3.06)**	(3.06)**	(2.52)**	(2.52)**
Heart problems	0.931	0.332	0.914	0.324	0.945	0.334
	(6.33)**	(6.33)**	(6.14)**	(6.14)**	(6.39)**	(6.39)**
Stroke	0.982	0.359	0.875	0.316	0.868	0.311
	(3.08)**	(3.08)**	(2.76)**	(2.76)**	(2.76)**	(2.76)**
Arthritis	0.719	0.248	0.448	0.146	0.686	0.233
	(5.47)**	(5.47)**	(3.18)	(3.18)	(5.17)**	(5.17)**
Emotional problems	0.764	0.264	0.842	0.293	0.717	0.243
	(6.35)**	(6.35)**	(6.92)	(6.92)	(5.87)**	(5.87)**
Pain	1.043	0.352				
	(11.75)**	(11.75)**				
Pain light			0.407	0.129		
			(3.72)**	(3.72)**		
Pain moderate			1.200	0.422		
			(11.08)**	(11.08)**		
Pain heavy			1.793	0.630		
			(9.49)	(9.49		
Neck pain					0.218	0.065
					(2.04)**	(2.04)**
Back pain					0.355	0.106
					(3.97)**	(3.97)**
Jaw pain					0.380	0.122
					(1.93)	(1.93)
Headache					0.077	0.022
					(0.77)	(0.77)
Joint pain					0.698	0.212
					(7.70)**	(7.70)**
Female	0.077	0.022	0.095	0.027	0.103	0.030
	(0.93)	(0.93)	(1.13)	(1.13)	(1.23)	(1.23)
Education med.	–0.057	–0.016	–0.103	–0.029	–0.091	–0.026
	(0.58)	(0.58)	(1.02)	(1.02)	(0.93)	(0.93)

(*continued*)

Table 9.7 (continued)

	Coefficient	DF/dX	Coefficient	DF/dX	Coefficient	DF/dX
Education high	–0.319	–0.089	–0.305	–0.084	–0.326	–0.089
	(3.16)**	(3.16)**	(2.98)**	(2.98)**	(3.21)**	(3.21)**
Age 35–44	–0.192	–0.053	–0.295	–0.079	–0.275	–0.073
	(1.33)	(1.33)	(2.02)**	(2.02)**	(1.89)	(1.89)
Age 45–54	0.030	0.009	–0.186	–0.051	–0.165	–0.045
	(0.22)	(0.22)	(1.33)	(1.33)	(1.17)	(1.17)
Age 55–64	0.174	0.052	0.127	0.037	0.140	0.041
	(1.20)	(1.20)	(0.88)	(0.88)	(0.97)	(0.97)
Age 65+	0.038	0.011	–0.114	–0.032	–0.092	–0.026
	(0.26)	(0.26)	(0.76)	(0.76)	(0.62)	(0.62)
Married	–0.114	–0.034	–0.147	–0.044	–0.106	–0.031
	(1.18)	(1.18)	(1.50)	(1.50)	(1.08)	(1.08)
Constant	–1.137		–1.100		–1.265	
	(6.88)**		(6.68)**		(7.55)**	
Observations	1537		1537		1537	
Observed *p*	0.254		0.254		0.254	
Log Likelihood	–643.50		–620.20		–635.99	
			United Kingdom			
High blood pressure	0.242	0.065	0.239	0.065	0.222	0.059
	(5.19)**	(5.19)**	(5.09)**	(5.09)**	(4.70)**	(4.70)**
Diabetes	0.441	0.131	0.480	0.146	0.456	0.136
	(4.65)**	(4.65)**	(5.06)**	(5.06)**	(4.74)	(4.74)
Cancer	0.977	0.335	0.962	0.330	0.960	0.327
	(7.00)**	(7.00)**	(6.85)**	(6.85)**	(6.76)**	(6.76)**
Heart problems	0.548	0.167	0.566	0.175	0.563	0.172
	(6.96)**	(6.96)**	(7.17)**	(7.17)**	(7.100)**	(7.100)**
Stroke	0.637	0.200	0.623	0.197	0.606	0.188
	(7.97)**	(7.97)**	(7.70)**	(7.70)**	(7.43)**	(7.43)**
Arthritis	0.641	0.193	0.627	0.190	0.568	0.168
	(13.57)**	(13.57)**	(13.11)**	(13.11)**	(11.83)**	(11.83)**
Emotional problems	0.660	0.206	0.663	0.208	0.620	0.191
	(10.89)**	(10.89)**	(10.93)**	(10.93)**	(10.00)**	(10.00)**
Pain	0.765	0.205	0.854	0.252		
	(21.21)**	(21.21)**	(22.75)**	(22.75)**		
Pain very mild					0.227	0.061
					(4.21)**	(4.21)**
Pain mild					0.461	0.133
					(8.19)**	(8.19)**
Pain moderate					0.873	0.272
					(17.56)**	(17.56)**
Pain severe					1.285	0.441
					(20.44)**	(20.44)**
Pain very severe					1.374	0.486
					(13.08)**	(13.08)**
Female	–0.049	–0.012	–0.057	–0.014	–0.070	–0.017
	(1.37)	(1.37)	(1.59)	(1.59)	(1.93)	(1.93)

Table 9.7 (continued)

	Coefficient	DF/dX	Coefficient	DF/dX	Coefficient	DF/dX
Education med.	–0.239	–0.058	–0.228	–0.056	–0.214	–0.052
	(5.86)**	(5.86)**	(5.58)**	(5.58)**	(5.19)**	(5.19)**
Education high	–0.235	–0.054	–0.218	–0.051	–0.192	–0.045
	(4.35)**	(4.35)**	(4.05)**	(4.05)**	(3.53)**	(3.53)**
Age 35–44	0.160	0.042	0.149	0.039	0.162	0.042
	(2.66)**	(2.66)**	(2.46)**	(2.46)**	(2.63)**	(2.63)**
Age 45–54	0.258	0.069	0.269	0.073	0.274	0.073
	(4.25)	(4.25)	(4.44)**	(4.44)**	(4.46)**	(4.46)**
Age 55–64	0.324	0.090	0.319	0.089	0.336	0.093
	(4.88)**	(4.88)**	(4.82)**	(4.82)**	(4.99)**	(4.99)**
Age 65+	0.499	0.140	0.508	0.144	0.520	0.146
	(7.73)**	(7.73)**	(7.89)**	(7.89)**	(7.92)**	(7.92)**
Married	–0.114	–0.029	–0.101	–0.026	–0.100	–0.025
	(2.87)**	(2.87)**	(2.53)**	(2.53)**	(2.48)**	(2.48)**
Constant	–1.538		–1.463		–1.624	
	(22.56)**		(21.79)**		(23.06)**	

Note: Robust *z* statistics in parentheses.

** Significant at the 1 percent level.

* Significant at the 5 percent level.

this goal, consider for example an evaluation of the impact of a single health condition *j*. Let $P(A)$ and $P(B)$ be the (predicted) work disability rates in country *A* and country *B* (for a given age group), and let $P(A)^{-j}$ and $P(B)^{-j}$ be the predicted work disabilities in countries *A* and *B* for the counterfactual situation that nobody would suffer from health problem *j*. We can then interpret $P(A) - P(A)^{-j}$ as the work disability rate in country *A* due to that health problem, and similarly for country *B*. Note that this assignment of importance to this health condition depends both on the prevalence of the health problem and on the sensitivity of the probability of work disability to that health problem (i.e., on the corresponding coefficients in β_A); we will separate these two below.

The difference in work disabilities in the two countries can be expressed using the following decomposition:

$$P(B) - P(A) = [P(B)^{-j} - P(A)^{-j}] + [P(B) - P(B)^{-j}] - [P(A) - P(A)^{-j}].$$

The first term on the right hand side can be interpreted as the difference between work disability prevalence in the two countries that is *not* due to the chosen health problem. The sum of the second and third term is then the part that is due to the chosen health condition. The latter two terms can be further separated in a *prevalence* effect (the percentage with the health problem) and an *impact* effect (the impact of the health problem on work disability). We can write:

$$P(A) - P(A)^{-j} = \frac{1}{N_A} \sum_{i \in A} [g(x_i, b_A) - g(x_i^{-j}, b_A)]$$

$$= \left(\sum_{i \in A} x_{ij} / N_A \right) \left[\sum_{i \in A, x_{ij}=1} \Delta g(x_j, b_A) / \sum_{i \in A} x_{ij} \right]$$

where $g(x_i, b_A)$ is the probability of having the health condition for an individual with characteristics x_i and parameter vector b_A.

The first factor is the fraction in country A that suffers from the chosen health problem (the quantity effect for country A). In the second term, $\Delta g(x_i, b_A)$ is the marginal effect (partial derivative) for a dummy variable, the difference if it is set to 1 or 0, with other variables set to their values for observation i. Thus, the second term can be seen as the average marginal effect for those who have the health problem.

The same decomposition can be used for all covariates in the model (both health and nonhealth dummy variables), allowing us to compare the importance of each to the reported rates of work disability in each country and the difference between the three countries.

Table 9.8 presents a summary of the relative contributions of different sets of factors toward explaining the differences between the three countries in reported rates of work disability. For this relative assessment, we divide covariates into five groups—the so-called objective health factors (hypertension, diabetes, cancer, diseases of the lung), heart problems and stroke, arthritis, emotional problems, and pain. The first three columns in table 9.8 assess the importance of each factor to explaining work disability in the Netherlands, the United States, and the United Kingdom. The final two columns assess the contribution of each factor toward explaining the differences between countries using the Netherlands as the reference group. Separate assessments are performed for each of the three models estimated for each country in table 9.7.

In each of the three countries, pain is by far the most important factor explaining reported rates of work disability. This is especially true for the Netherlands and the United Kingdom, where observed work disability rates are higher than in the United States. Moreover, as summarized by the "all pain" row in table 9.8, the estimated role of pain rises when we estimate models which differentiate between the degree of pain (light, moderate, and heavy) and the location of pain in the body. Joint pain, and to a somewhat lesser degree back pain, are the most central types of pain in explaining rates of work disability.

The most important columns in table 9.8 are the final two which summarize the role of each set of factors toward explaining differences in work disability between the countries. Once again compared to either the Netherlands or the United Kingdom, pain predicts much lower rates of work disability in the United States. This is in part due to the lower pain

Table 9.8 Contributions of factors to explaining work disability

	Netherlands	United Kingdom	United States	Netherlands–U.K.	Netherlands–U.S.
		Model 1			
Objective health	1.57	2.17	2.64	–0.60	–1.07
Heart problems	2.38	1.61	1.76	0.77	0.62
Arthritis	2.34	2.86	2.74	–0.52	–0.40
Emotional	2.44	1.30	1.72	1.14	0.72
Pain	8.50	6.63	3.05	1.87	5.45
		Model 2			
Objective health	1.48	2.03	2.52	–0.55	–1.04
Heart problems	2.15	1.57	1.78	0.58	0.37
Arthritis	1.34	2.59	2.19	–1.25	–0.85
Emotional	2.61	1.19	1.68	1.42	0.93
Pain light	1.48	2.08	0.05	–0.60	1.43
Pain moderate	6.37	3.98	1.40	2.39	4.97
Pain heavy	3.19	3.82	1.22	–0.63	1.97
All pain (sum of above three rows)	11.04	9.88	2.67	1.16	8.37
		Model 3			
Objective health	1.60		2.40		–0.80
Heart problems	2.31		1.70		0.61
Arthritis	2.29		1.91		0.38
Emotional	2.25		1.54		0.71
Back pain	3.45		1.72		1.73
Joint pain	7.88		2.13		5.75
Other Pain	2.28		1.27		1.01
All pain (sum of above three rows)	13.61		5.12		8.49

Note: Units are percentage points.

prevalence in the United States and in part due to the lower effect of pain on work disability in the United States compared to the other two countries. In explaining lower rates of work disability in the United States, pain is by far the most important factor of those listed in table 9.8. Why individuals in the United States respond less to pain than residents of the other two countries will be the central question in the next two sections.

9.4 Pain Medication and Workplace Accommodation

How pain translates into a personal assessment of a work disability may be affected by pain medication and the types of accommodations available in the workplace to deal with any impairment. If pain medication alone sufficiently alleviates the symptoms and severity of the pain, individuals

may not feel that they actually have a work disability. Similarly, if accommodations are available at work so that the impairment does not affect the daily routines of work or how productive a worker is, individuals may also believe that their problems are not relevant to their current work situation. In both situations, individuals may answer a question on whether they have a work disability in the negative even though without medication or accommodation they would have one. Moreover, both the use and availability of pain medication or the extent of accommodations available at work may well vary across the three countries we are studying. If they do, these two factors may account for some of the differences in reported work disability across these countries. To investigate this possibility, we present information in this section on the role of pain medication and workplace accommodation in each of our three countries.

9.4.1 Pain Medication

To help answer these questions, we added a pain module to the December 2004 wave of the Dutch CentERpanel. To the question on whether they were often troubled by pain, respondents could answer (1) yes, (2) no, because I use pain medication, and (3) no, and I do not need pain medication. If people respond yes, there was a follow-up question that inquired about whether they used pain medication to combat the pain. That sequence of questions allows us to estimate how many people troubled by pain are using pain medication and how effective that medication is in eliminating the pain.

The results are listed in table 9.9. The use of pain medication is actually very widespread in the Netherlands and the use of this medication affects the reporting of pain. While 26.5 percent of respondents reported that they were often troubled with pain, that fraction would grow to 37.4 percent if we included those whose pain medication eliminated the pain. Among the Dutch respondents who either had pain or would have had pain without medication, 69 percent were taking medication for this pain. Moreover, the use of this medication was quite effective. Within this group, 42 percent of Dutch respondents had no pain at all. Using this definition of effectiveness, pain medication appears equally effective for women and men, but appears to have eliminated pain completely in a larger fraction of the more educated Dutch respondents. This may be due to the fact that their pain was less severe.

Unfortunately, the pain medication questions in the United States and the United Kingdom are not strictly comparable to those in the Netherlands. For the United States we use data from the National Health and Nutrition Examination Survey (NHANES), which asked similar questions about the location of pain (neck, back, headaches, joint, face) during the last three months as described previously for the NHIS. The advantage of

Table 9.9 The use of pain medication

	All	Men	Women	Ed low	Ed med	Ed high	Ages 45+	Ages 45–64
			I. The Netherlands					
A. % with pain or taking painkillers	37.4	28.9	46.5	40.8		29.8	41.9	42.9
B. % of A taking painkillers	68.9	64.7	71.6	69.4		70.6	66.4	67.4
C. % of B with no pain	41.6	43.9	40.9	39.1		49.1	36.9	37.4
% with pain	26.5	20.7	33.1	29.9		19.5	31.6	32.1
			II. United States					
A. % with pain or taking painkillers	61.6	57.1	65.7	64.1	65.1	58.7	65.7	64.3
B. % of A taking painkillers	41.3	41.0	41.5	37.3	43.8	42.1	54.7	48.4
C. % of B with no pain	35.5	43.9	29.1	30.6	29.9	41.1	40.3	38.2
% with pain	52.6						51.3	52.4
			III. United Kingdom					
	All 52+							
A. % with pain	38.3	33.7	41.9	41.9		30.1		
B. % with moderate/severe pain	27.7	25.7	29.0	28.7		24.4		
C. % of B taking pain medication	27.3	21.2	31.0	26.6		29.7		
D. % of B with pain being controlled	60.1	53.2	62.9	59.2		63.1		

Source: Netherlands—CentERpanel, December 2004. United States—NHANES 1999–2000. Pain is defined as some form of pain in the last three months, including neck, face, back, headaches, or joint pain. United Kingdom—ELSA 2004. Sample is aged fifty in 2002.

NHANES is that it also contains a detailed set of questions about all types of medications. The noncomparability with the Dutch sample derives from the fact that we have already demonstrated that this form of the pain question elicits much higher prevalence rates than the recurrent pain question. This expansion in pain prevalence no doubt includes many less serious forms of pain.

For the United Kingdom we use new data from the latest wave of the English Longitudinal Study of Ageing (ELSA), which contains detailed questions on certain types of pain alleviation as part of their questions on the use and efficacy of health care services. In this case the noncomparability arises for three reasons. First, only individuals reporting moderate or severe pain are asked general questions about pain medication. Second, for both general and specific types of pain medication, the ELSA questions relate solely to medication or treatment prescribed by a respondent's doctor or nurse. Finally, the ELSA sample consists of individuals aged fifty and over in 2002, as opposed to being an age-representative sample such as the NHANES or CentERpanel.

These important caveats should be kept in mind when interpreting the second and third panels of table 9.9, illustrating the extent of pain medication in the United States and the United Kingdom respectively. Among

those with pain or without the symptoms of pain due to medication, a much smaller proportion of Americans (41.3 percent) are taking pain medication. When they do take medication, it also appears to be less effective in completely eliminating pain symptoms than it was for Dutch respondents. In the United Kingdom, an even lower fraction report receiving medication than in the United States (even when the definition of pain medication in the United States is limited to prescription painkillers only). This effect may even be somewhat underestimated since those in mild pain (who presumably have an even lower rate of medication) are routed out of the ELSA questions. On the other hand, those receiving medication are much more likely than those in both the United States and the Netherlands to report that the medication controls their pain. Once again, comparability of question wording may be an issue here. If controlled pain equates to mild pain, then such cases will be differentially recorded across the different surveys.

Despite the relative lack of comparability of these data, the relevance of their overall message to the questions addressed in this chapter is clear. While we observe a much lower prevalence of work disability and pain in the United States and the United Kingdom compared to the levels observed in the Netherlands, it is not due to a higher rate of (successful) medication in the United States and the United Kingdom. If anything, the differences across countries appear to go the other way.

9.4.2 Workplace Accommodation

In December 2004, we fielded a module on work disability in the Dutch CentERpanel that was based on one already used in ELSA. This module posed a series of questions on workplace accommodations to all respondents who were not self-employed and who had worked during the last decade. These respondents were asked if they had ever asked their employer to make an accommodation, whether their employer had ever offered to make an accommodation, and whether their employer had ever made an accommodation. The types of accommodation inquired about included making work less physically demanding, less mentally demanding/stressful, reducing hours worked/arranging job-sharing, making working hours more flexible, allowing work from home, and providing special equipment and other such adaptations to the workplace that make it easier to keep working.

A unique aspect of this module is that this series of questions were asked of all respondents, whether or not they currently have a work disability. As will be the case with the American and British survey on workplace accommodations discussed below, the standard practice is to restrict these questions to those who said that they had a workplace disability. The advantage of the protocol used in Dutch panels is that it provides a complete description of the availability of workplace accommodations in the work

force. For example, if the provision of effective workplace accommodations induced some respondents to say that they did not have a workplace disability, we would never be able to know that with questions limited to those with a workplace disability.

Tables 9.10 and 9.11 summarize the responses from the Dutch respondents from the work accommodation module. Table 9.10 provides the data on the full set of respondents, while table 9.11 is limited to the subset that reports that they have a work disability.

There are no salient differences by age in these patterns of workplace ac-

Table 9.10 Dutch answers on work accommodation for full sample (answers in percents)

Variable	Age 25+	Age 45–64	Age 45+	Men	Women	Low Education	Hign Education
Currently employed	54.8	53.1	35.1	58.9	50.5	52.4	60.2
Ever employed	94.1	94.3	94.2	97.9	90.7	93.4	95.8
Ever asked employer to change job to							
Less physically demanding	15.6	17.0	15.8	15.7	15.7	19.3	8.2
Less stressful	20.8	21.8	20.9	19.3	22.5	20.7	20.9
Reduce hours	19.2	20.6	19.6	15.6	23.7	18.3	21.1
Make hours flexible	16.8	15.8	15.5	16.3	17.3	15.7	19.0
Work from home	14.2	12.8	12.0	15.6	12.5	11.4	19.8
Provide special equipment	26.1	24.1	22.9	24.4	28.1	28.7	20.9
Other	9.9	12.0	11.3	10.6	9.2	11.2	7.3
Employer ever offered to change job to							
Less physically demanding	17.1	17.0	16.2	16.8	17.4	20.2	10.8
Less stressful	16.0	15.7	15.1	14.5	17.8	16.9	14.1
Reduce hours	13.4	14.3	14.9	12.4	14.6	13.4	13.5
Make hours flexible	16.9	16.0	16.0	17.7	15.9	16.4	18.0
Work from home	11.9	11.7	11.2	13.1	10.3	8.3	19.0
Provide special equipment	26.6	24.0	23.1	25.6	27.9	29.1	21.5
Other	5.0	3.8	3.5	5.4	4.7	5.9	3.2
Employer ever changed jobs to							
Less physically demanding	15.1	14.9	14.4	14.4	15.8	18.2	8.8
Less stressful	11.9	12.8	12.4	9.6	14.7	12.4	10.8
Reduce hours	15.5	15.9	16.2	13.2	18.4	14.3	18.1
Make hours flexible	17.0	16.6	17.2	16.2	18.0	15.8	19.5
Work from home	9.7	10.6	10.0	10.0	9.2	5.7	17.6
Provide special equipment	25.3	22.1	21.8	22.9	28.2	27.3	21.3
Other	3.0	2.2	2.4	2.8	3.2	3.3	2.3
Had adjustment helped	86.2	82.8	83.2	83.8	88.7	86.6	85.4
Would adjustment have helped	23.6	23.3	21.9	23.8	23.2	22.7	25.4

Notes: "Ever Employed": only asked to those who are not current employees. "Physically Demanding, . . . , Other": only asked to current employees and those who have been employees ever since 1996. "Has Adjustment Helped": only asked to those for whom at least one actual adjustment was made. "Would Adjustment Have Helped": only asked to those for whom no adjustments were made.

Table 9.11 **Dutch answers on work accommodation for those with current work disability (answers in percents)**

Variable	Age 25+	Age 45–64	Age 45+	Men	Women	Low Education	Hign Education
Currently employed	33.5	30.9	20.6	34.8	32.4	32.6	36.1
Ever employed	94.9	96.4	95.1	97.9	92.6	95.8	91.9
Ever asked employer to change job to							
Less physically demanding	35.0	34.0	31.9	38.4	31.5	40.2	19.8
Less stressful	30.2	30.1	29.9	28.7	31.8	30.6	29.2
Reduce hours	32.1	33.7	32.9	22.5	25.8	23.8	25.2
Make hours flexible	24.1	24.4	24.4	16.3	17.3	15.7	19.0
Work from home	16.6	10.6	9.8	16.5	16.6	16.9	15.6
Provide special equipment	36.0	32.8	32.1	32.8	39.2	37.5	31.4
Other	18.4	22.7	21.5	22.1	14.7	18.3	18.8
Employer ever offered to change job to							
Less physically demanding	28.4	26.4	25.3	30.7	26.1	32.3	17.0
Less stressful	21.9	20.5	20.2	20.8	23.0	24.8	13.4
Reduce hours	24.1	23.6	23.5	26.2	21.9	24.4	23.0
Make hours flexible	21.8	19.4	19.2	22.6	21.1	22.6	19.6
Work from home	11.0	9.8	9.4	11.4	10.5	10.7	11.6
Provide special equipment	30.2	27.6	27.4	25.5	34.9	32.7	22.8
Other	7.3	3.9	3.9	8.0	6.5	7.0	7.9
Employer ever changed job to							
Less physically demanding	28.0	23.9	23.4	30.6	25.3	32.1	15.9
Less stressful	17.1	18.5	18.0	14.3	19.7	19.0	11.1
Reduce hours	24.9	24.0	24.3	25.6	24.1	25.1	24.2
Make hours flexible	23.1	17.4	18.2	21.6	24.7	23.9	20.8
Work from home	7.8	8.2	7.7	8.4	7.3	7.1	10.1
Provide special equipment	29.7	26.9	27.1	25.0	34.5	33.1	20.0
Other	5.0	3.3	3.7	4.7	5.4	4.0	8.2
Had adjustment helped	78.3	73.7	74.2	74.5	82.4	77.5	81.1
Would adjustment have helped	34.3	31.8	0.0	36.7	32.0	31.9	40.6

Notes: "Ever Employed": only asked to those who are not current employees. "Physically Demanding, . . . , Other": only asked to current employees and those who have been employees ever since 1996. "Has Adjustment Helped": only asked to those for whom at least one actual adjustment was made. "Would Adjustment Have Helped": only asked to those for whom no adjustments were made.

commodations. The principal differences that emerge by gender have to do with flexibility of hours where women are more likely to ask and to have had adjustments in their work hours. However, this pattern is only apparent in the full sample, which suggests that the differential gender treatment is largely due to other matters (such as family responsibilities) rather than work disabilities. Within the work disabled subsample, women are more likely to have had adjustments in their physical workplace while men are more likely to have equipment adjustments.

There are much stronger differences by education that appear in both

the full and work disability samples. Those in the lower education category are much more likely to have asked for, been offered, and received physical and equipment adjustments in their workplace environment. For example, among those with a work disability, 32 percent of less educated Dutch respondents had a physical adjustment to their workplace compared to only 16 percent of the higher educated respondents.

The final two rows in these tables provide a summary of the Dutch respondents assessment of whether these workplace accommodations were helpful. When there were workplace accommodations, more than three quarters of respondents thought that the adjustments were useful, and when there were no workplace adjustments a third of respondents still believed that the adjustments would have helped if they had been made.

As previously explained, questions on workplace accommodations in American surveys are limited to those with a work disability. Perhaps, the best module was placed into the HRS, where a set of questions was asked about workplace accomodations for those with a work disability. These questions were asked whether or not the individual was currently employed. If not currently employed, the questions referred to the last time of employment.

Table 9.12 (based on the HRS) provides a description of the types of help provided by employers. These data in the HRS sample are most comparable with data from the Dutch samples that are restricted to those with a current work disability and who are older workers (forty-five to sixty-four in the Dutch sample). Similar to the Dutch case, gender differences in workplace accommodation in the United States are small. But in sharp contrast to the Dutch data, there is also almost no education gradient to the use of workplace accommodation in the United States. Most importantly, workplace accommodations are far less common in the American than in the Dutch workplace. This generalization appears to be true across the board,

Table 9.12 **Workplace accommodation in the United States (answers in percents)**

	All	Men	Women	Low ed	Med. ed	High ed
Did employer help you	22.4	22.1	22.7	22.4	21.3	24.5
Somewhat helped you out	9.3	8.7	9.5	9.6	9.5	6.9
Shorter work day	6.3	6.4	6.3	6.1	5.8	9.0
Flexible hours	7.3	6.6	7.9	8.6	7.2	9.8
More breaks	8.5	8.5	8.5	8.6	6.5	8.2
Special transportation	1.2	0.9	1.4	1.3	1.3	0.4
Change job	10.1	11.3	8.9	10.2	9.2	11.0
Help learn new skills	3.1	2.5	3.7	3.2	3.1	2.4
Special equipment	2.7	2.6	2.8	2.7	2.8	3.3
Anything else	6.4	6.1	6.7	6.2	6.0	8.6

Note: 1992-HRS baseline ages fifty-one to sixty-one. Sample: all those who said that they had a work disability.

but it is especially pronounced for equipment and physical changes in the workplace.

Since the workplace accommodation questions for our ELSA sample are limited to those who are currently employed, table 9.13 contains the most directly comparable data for all three countries. In this table, both the Dutch and American data are also limited to those who are currently employed. In addition, to preserve some age comparability, the Dutch sample is limited to those forty-five to sixty-four and the American sample to those ages fifty-one to sixty-one. While this is the most comparable comparison possible between all three countries, it is important to note that sample sizes in the Dutch sample become quite small.

The first panel of table 9.13 summarizes the responses from the Dutch respondents from the work accommodation module. To enhance comparability across surveys, we select the sample of older respondents who report a work disability but who were also working at the time of the survey. The principal differences that emerge by gender have to do with the physical nature of work, where women are less likely to have had adjustments, and in flexibility of hours and special equipment, where women are more likely to have had adjustments. Differences by education are also apparent. As before, those in the lower education category are much more likely to have asked for, to have been offered, and to have received physical and equipment adjustments in their workplace environment.

The 2004 wave of the English Longitudinal Study of Ageing contains the same questions on workplace accommodation, although due to the design of the survey, some individuals are routed out of some of the items. In table 9.13 we show similar descriptive statistics to those from the Netherlands for the ELSA sample (which is aged fifty-two and over in 2004). The first three lines of this table establish some basic patterns in the data. As observed in earlier sections of this chapter the prevalence of work disability is high, and higher amongst the low education group than the high education group. In addition, conditional on reporting a work disability, the high education group is substantially more likely to work, but conditional on having a work disability and being in work, the two education groups are equally likely to report that their work disability limits their activities in the current job.

What is apparent from the across-country comparison in table 9.13 is that both overall levels and the patterns across accomodations and across gender and education subgroups are quite different in the United Kingdom from those observed in the Netherlands. Individuals working with a work disability in the United Kingdom are much less likely to have received modifications to their work environment in the United Kingdom. The overall level of accommodations is twice as high in the Netherlands as in the United Kingdom, and the differences are even greater when looking at each individual type of accommodation separately. Perhaps more surprisingly,

Table 9.13 Workplace accommodation of disability

	All	Men	Women	Low Education	Med. Education	High Education
			I. Netherlands			
Did employer help you in any way	70.6	77.9	58.5	75.4		59.5
Physically less demanding	28.3	37.2	13.5	34.2		14.8
Less stress	25.1	26.0	23.6	29.2		15.9
Shorter work day	26.5	27.4	25.0	25.1		29.7
Flexible hours	18.4	16.5	21.6	20.0		14.7
Work from home	10.3	14.3	3.5	7.7		15.9
Special equipment or adjustment	33.2	26.4	44.6	34.3		30.9
Anything else	6.3	6.4	6.3	4.6		10.2
			II. United Kingdom			
A. Percent of those aged 52+ reporting a work disability	33.1	33.0	33.2	36.5		25.3
B. Percent of A who are working	13.3	14.4	12.5	10.4		22.9
C. Percent of B whose work disability limits type or amount of work in current job	42.9	41.2	44.5	41.9		44.4
D. All employees reporting a work disability						
Percent whose employer has either changed or offered to change their work to make it:						
Less physically demanding	9.9	12.3	8.0	9.8		10.0
Less mentally demanding/stressful	2.5	1.6	3.1	2.3		2.7
Fewer hours/job sharing	5.6	4.1	6.8	4.0		8.2
More flexible hours	3.5	2.5	4.3	2.9		4.5
Working from home sometimes	1.8	0.8	2.5	0.6		3.6
Special equipment/workplace adaptation	8.1	5.7	9.9	5.7		11.8
Other	2.1	0.0	3.7	1.7		2.7
Any of the above	25.7	22.1	28.4	21.3		32.7
			III. United States			
Did employer help you	29.6	28.4	31.2	32.6	26.0	21.8
Somewhat helped you out	11.6	8.9	15.4	13.8	8.5	6.5
Shorter work day	8.3	8.9	7.5	10.0	3.9	6.9
Flexible hours	10.1	9.9	10.5	12.9	4.0	5.4
More breaks	11.5	11.5	11.6	13.8	6.8	8.2
Special transportation	1.5	0.9	2.3	1.7	0.9	1.4
Change job	16.5	17.4	14.8	19.2	10.5	12.4
Help learn new skills	4.6	4.8	4.3	5.1	4.8	2.2
Special equipment	4.4	5.5	2.8	5.3	3.2	2.3
Anything else	6.8	5.8	8.2	6.8	5.3	8.1

Notes: The Netherlands—2004 CentERpanel ages 45–64. Sample: all those who said that they had a work disability and who were at work at the time of the survey (ninety-one observations). United Kingdom—2004 ELSA data ages fifty-two and over—sample all those who said that they had a work disability and who were at work at the time of the survey. United States—1992 HRS baseline ages fifty-one to sixty-one. Sample all those who said that they had a work disability and who were at work at the time of the survey. Data are weighted.

the differences by gender and education are reversed. In the United Kingdom it is women and the highly educated who are most likely to have received workplace accommodations (conditional on working), whereas in the Netherlands these groups have a lower likelihood of workplace accommodation. Once again, evidence from the United States, presented in panel C of table 9.13, reveals similarities between the United States and the United Kingdom and differences to the Netherlands. Table 9.11 (based on the HRS baseline data) provides a description of the types of help provided by employers. The overall level of employer accommodation is lower even than in the United Kingdom (although it should be remembered that the HRS baseline data was collected in 1992, some twelve years before the ELSA data presented for the United Kingdom). As in the United Kingdom, women are more likely to receive accommodations, but as in the Netherlands, it is the more educated that are more likely to receive workplace accommodations in the United States.

This section began by offering the possibility that some of the difference in work disability prevalence among these three countries was due to differences in the use of either pain medications or workplace accommodations. If the use of pain medications or workplace accommodations was more common in the United States that could partially explain the lower rates of reported work disability there. However, if anything, the patterns go the other way with less frequent use of work accommodations and medication in the United States. Apparently, explanations for lower reported rates of work disability in the United States must lie elsewhere.

9.5 Vignettes

If differential use of pain medication and workplace accommodation across countries cannot explain cross country differences in work disability prevalence as documented in section 9.3, what may explain it? In this section we present and apply a new methodology that aims at uncovering differences across countries in their norms and attitudes toward work disability. This new methodology relies on the use of vignettes.

We first provide an intuitive description of the use of vignettes for identifying reporting biases, following King et al. (2004) and Kapteyn, Smith, and Van Soest (2007). Their model shows how vignettes can help to identify systematic differences in response scales between groups (or countries), making it possible to decompose observed differences in, for example, self-reported health in a specific domain into differences due to response scale variation and genuine differences in health. Our analysis applies this model to work limiting disability rather than health. Vignette evaluations were collected in the Netherlands in the fall of 2003, and in the United States in early 2004. Work disability vignettes for the United King-

dom are not available yet. Thus, we can only compare the United States and the Netherlands.

9.5.1. Using Vignettes to Identify Response Scales in Pain

The basic idea of the model is sketched in figure 9.1. It presents the distribution of work-related health in two countries. The density of the continuous health variable in country A is to the left of that in country B, implying that on average, people in country A are less healthy than in country B. The people in the two countries, however, use different response scales if asked to report their health on a five-point scale (for example, poor, fair, good, very good, excellent). In our example, people in country A have a much more positive view on a given health status than people in country B. For example, someone in country A with the health indicated by the dashed line would report to be in very good health, while a person in country B with the same actual health would report fair. The frequency distribution of the self-reports in the two countries would suggest that people in country A are healthier than those in country B—the opposite of the actual health distribution. Correcting for the differences in the response scales, differential item functioning (DIF)—in the terminology of King et al. (2004)—is essential to compare the actual health distributions in the two countries.

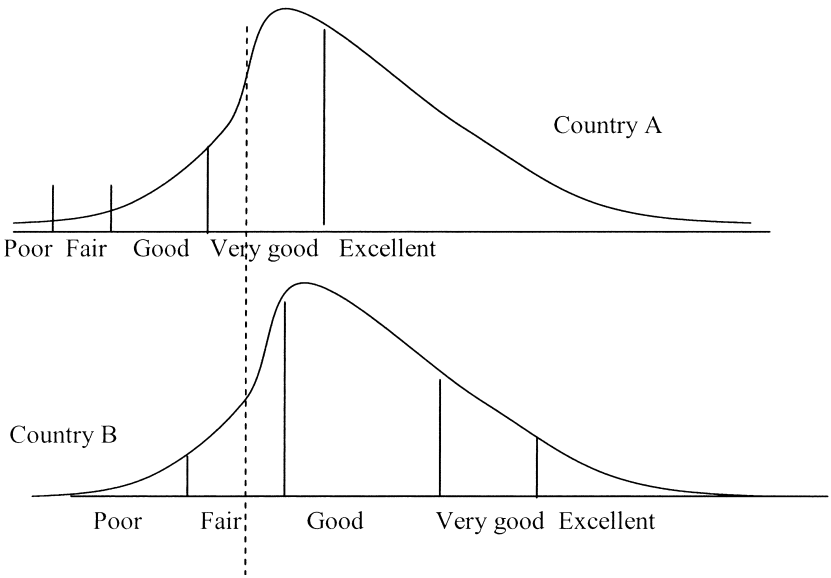

Fig. 9.1 Comparing self-reported health across two countries in case of DIF

Vignettes can be used to do the correction. A vignette question describes the health of a hypothetical person and then asks the respondent to evaluate that person's health on the same five-point scale that was used for the self-report. For example, respondents can be asked to evaluate the health of a person whose health is given by the dashed line. In country A, this will be evaluated as very good. In country B, the evaluation would be fair. Since the actual health description of the vignette person is the same in the two countries, the difference in the evaluations must be due to DIF. Vignette evaluations thus help to identify the differences between the response scales in the two countries. 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 underlying assumption is *response consistency*: a given respondent uses the same scale for the self-reports and the vignette evaluations.

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 figure 9.1, this will lead to the correct conclusion that people in country B are healthier than those in country A, on average. King et al. (2004) develop parametric and nonparametric models that make it possible to perform the correction. They apply their method to, for example, political efficacy and visual acuity. Their results strongly support the ability of the vignettes to correct for DIF. For example, in a comparative study of political efficacy of Chinese and Mexican citizens, they find that without correction the Chinese seem to have more political influence than the Mexicans. The conclusion reverses if the correction is applied.[6]

9.5.2 Econometric Model

The model explains respondents' self-reports on work limitations and their reports on work limitations of hypothetical vignette persons. The first is the answer (Y_{ri}, i respondent i) to the question:

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

In our data for the United States, the answers are given on a yes/no scale. In the Dutch data, respondents answer this question both on a yes/no scale and on a five points scale, with answers no, not at all ($Y_{ri} = 1$), yes, I am somewhat limited ($Y_{ri} = 2$), yes, I am moderately limited ($Y_{ri} = 3$), yes, I am very limited ($Y_{ri} = 4$) and yes, I am so seriously limited that I am not able to work ($Y_{ri} = 5$).

Table 9.2 suggests that there is some random error in the two-point and/or five-point scale evaluations that is not transferred to the other scale. To account for this, we use the following equations for the respondent's own

6. More applications to health are discussed in Salomon, Tandon, and Murray (2004).

work limiting disability, partitioning the error term in a genuine unobserved component of work disability affecting both the two-point and the five-point scale reports, and an idiosyncratic error term affecting only one scale and independent of everything else:

Genuine work disability:

$$Y^*_{ri} = X_i\beta + \varepsilon_{ri};\ \varepsilon_{ri} \quad N(0, \sigma_r^2),\ \varepsilon_{ri} \text{ independent of } X_i,\ V_i.$$

Five-point scale self-reports:

$$Y_{ri} = j \text{ if } \tau_i^{j-1} < Y^*_{ri} + u_i^5 \leq \tau_i^j,\ j = 1, \ldots, 5.$$

Two-point scale self-reports:

$$Y_{ri} = 0 \text{ if } Y^*_{ri} + u_i^2 \leq \tau_i(2)$$

$$Y_{ri} = 1 \text{ if } Y^*_{ri} + u_i^2 > \tau_i(2)$$

$$u_i^2, \quad N(0, \sigma^2_{u^2});\ u_i^5 \quad N(0, \sigma^5_{u^5});\ u_i^2, u_i^5 \text{ independent of each other and of other errors (such as } \varepsilon_{ri}).$$

The thresholds τ_j^i between the categories of the five-point scale are given by:

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

The fact that different respondents can use different response scales is called differential item functioning (DIF). As in the King et al. (2004) model, we assume that response scales can vary only with observed characteristics V_i, including a country dummy and interactions with that country dummy. The exponentials guarantee that the thresholds increase with j.

In order to link the two-point scale and the five-point scale, we use the fact that the cutoff point between *yes* and *no* for the two-point scale is somewhere between the cutoff points between *no* and *mildly* and *mildly* and *moderately* for the five-point scale. In line with this, we model the cutoff point $\tau_i(2)$ on the two-point scale as a weighted mean of the two first cutoff points on the five-point scale:

$$\tau_i(2) = \lambda\tau_i^1 + (1 - \lambda)\tau_i^2$$

We assume that the weight λ does not vary with individual characteristics and is the same in the United States and the Netherlands. Thus, the thresholds on the five-point scale and the thresholds on the two-point scale can have completely different structures in the two countries, but the relation between them is the same. If the Dutch have lower thresholds on the five-point scale, they also have a lower threshold on the two-point scale, etc. This assumption is needed as long as there are no five-point scale self-reports on the five-point scale for the United States. Intuitively, it seems clear that the parameter λ can be identified from the Dutch self-reports on both scales.

In the United States as well as the Netherlands, the questions on work limitations of the vignette persons have the same five answering categories as the five-point scale self-report, and are formulated in the same way ("Does Mr./Mrs. X have any impairment or health problem that limits the type or amount of work that he or she can do?"). The answers will be denoted by Y_{li} where each respondent i evaluates a number of vignettes $l = 1, \ldots, L$.

The evaluations of vignettes $l = 1, \ldots, L$ are modeled using a similar ordered response model:

$$Y^*_{li} = \theta_l + \theta \text{ Female}_{li} + \varepsilon_{li}$$

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

$$\varepsilon_{li} \sim N(0, \sigma^2), \text{ independent of each other, of } \varepsilon_{ri} \text{ and of } X_i, V_i.$$

An important assumption is that the thresholds τ_i^j are the same for the five-point self-reports and the vignettes (*response consistency*). This is the basis for why vignettes help to identify DIF and help to correct for reporting differences.

The second assumption of King et al. (2004) is that Y^*_{li} does not vary with respondent attributes in any systematic way; it only varies with vignette characteristics given in the descriptions of the vignettes (captured by a vignette specific constant θ_i and a dummy for the gender of the vignette person).

Given these assumptions, vignette evaluations can be used to identify β and γ ($= \gamma^1, \ldots, \gamma^5$) if all questions were asked on the five-point scale: from the vignette evaluations alone, γ, θ, $\theta_1, \ldots, \theta_5$ 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, which is asymptotically efficient.

Correcting for DIF is straightforward once the parameters are estimated. Define a benchmark respondent with characteristics $V_i = V(B)$. (For example, choose one of the countries as the benchmark country.) The DIF correction would now involve comparing Y^*_{ri} to the thresholds τ^j_B rather than τ^j_i, where τ^j_B is obtained in the same way as τ^j_i but using $V(B)$ instead of V_i. Thus, a respondent's work-related health is computed using the benchmark scale instead of the respondent's own scale. This does not lead to a corrected score for each individual respondent (since Y^*_{ri} is not observed) but it can simulate corrected *distributions* of Y_{ri} for the whole population or conditional upon some of the characteristics in V_i and or X_i. Of course, the corrected distribution will depend upon the chosen benchmark.

9.5.3 Data and Vignette Questions

To estimate the model comparing work disability in the United States and the Netherlands, three data sets are combined: the Dutch CentER-panel (waves 1, 2, and 3 in August, October, and December 2003), the US RAND MS Internet panel, and the US HRS wave 1. They all have different age selections (all age groups in CentERpanel, 40+ in RAND MS Internet Panel, fifty-one to sixty-one in HRS), but since we condition on age, this should not be a problem. CentERpanel and RAND MS have exactly the same vignette questions on pain problems, emotional problems, and cardiovascular disease. HRS wave 1 has no vignettes. In this chapter, we only use the vignettes on pain problems.

In August 2003, we have collected work disability self-reports and vignette evaluations in the Dutch CentERpanel, which allows researchers to include short modules of experimental questions. This feature has been used to collect our data on work disability. The Internet infrastructure makes the CentERpanel an extremely valuable tool to conduct experiments, with possibilities for randomization of content, wording, question and response order, and regular revisions of the design. Production lags are very short, with less than a month between module design and data delivery. Based upon our first analysis, we have fielded a second wave in October with different wordings of the vignette questions. In this chapter, we use the self-reports on work disability collected in the first wave (August 2003) and we use vignette data from both waves (August and October 2003). The vignettes on pain are presented in table 9.14. All of them deal with back pain. The first two describe relatively light problems; the other three describe more serious problems.

The vignette questions in table 9.14 were also fielded in the RAND MS Internet panel, an Internet survey for U.S. respondents aged forty and over. Table 9.15 presents the vignette evaluations in the United States and the Netherlands. In both countries, the frequency distributions of evaluations reflect that vignettes 1 and 2 describe less serious problems than vignettes 3, 4, and 5. Still, there are some substantial differences in the evaluations between the two countries. In particular, for the first two vignettes, the U.S. respondents much more often report that the described persons have no limitation at all, where the Dutch respondents have a larger tendency to use the intermediate categories mildly and moderately. The same tendency towards the extremes in the United States and towards the middle for the Netherlands is seen in the fourth vignette, describing a person with relatively serious work limitations. The U.S. respondents much more often evaluate this person as severely or extremely limited, where the Dutch still tend to use the answer moderately. This suggests that correcting for response scale differences could reduce the difference in self-reported health distributions between the two countries.

Table 9.14 Vignette descriptions on pain problems

1. [Katie] occasionally feels back pain at work, but this has not happened for the last several months now. If she feels back pain, it typically lasts only for a few days.
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. [Yvonne] has almost constant pain in her back and this sometimes prevents her from doing her work.
4. [Jim] has back pain that makes changes in body position while he is working very uncomfortable. He is unable to stand or sit for more than half an hour. Medicines decrease the pain a little, but it is there all the time and interferes with his ability to carry out even day-to-day tasks at work.
5. [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.

Table 9.15 Vignette evaluations in the United States and Netherlands (percent distributions)

	Vignette 1		Vignette 2		Vignette 3		Vignette 4		Vignette 5	
Limited?	NL	U.S.	NL	U.S.	NL	U.S.	NL	U.S.	NL	U.S.
Not at all	24.89	38.09	10.52	29.66	0.35	0.15	0.46	0.15	0.46	0.73
Mildly	63.28	49.71	53.46	47.87	6.22	7.35	7.28	2.35	11.94	8.50
Moderately	10.47	10.44	29.44	20.26	26.56	30.44	31.11	15.42	33.79	38.56
Severely	1.32	0.88	6.27	1.47	50.89	46.76	46.28	58.88	43.90	40.91
Extremely	0.05	0.88	0.30	0.73	15.98	15.29	14.87	23.20	9.91	11.29

Sources: Netherlands—CentERpanel, August 2003, 1,977 observations; United States—RAND MS Internet Panel, 2003–2004, 681 observations.

9.5.4 Estimation Results

Estimation results of the complete model are presented in table 9.16. The equations for work disability and for the thresholds all include a complete set of interactions with the country dummy for the Netherlands. Vignette evaluation equations and the auxiliary parameters introduced above concerning the transformation from the two-point to the five-point scale do not include such interactions. Panel A of table 9.16 presents the results for the work disability equation in the complete model and in a model without any form of DIF, in which thresholds do not vary by country, individual characteristics, or health conditions. The latter model is clearly rejected against the complete model by a likelihood ratio test.

Education level in the United States is more important according to the complete model than in the model without DIF. The explanation is that the pain vignettes indicate that in the U.S., the higher educated use lower thresholds than the lower educated (i.e., they tend to assign higher work disability to the same vignette person than the lower educated). This is also

Table 9.16 **Estimation results United States—Netherlands model**

Panel A[a] Work disability

	Model without DIF		Complete model	
	est.	s.e	est.	s.e.
Constant	–10.424	1.444*	–11.033	1.560*
Education med.	–2.425	0.346*	–3.294	0.584*
Education high	–4.857	0.509*	–5.933	0.809*
Age 15–44	–17.359	6.287*	–15.996	8.365+
Age 45–54	–2.740	1.345*	–1.665	1.620
Age 55–64	–0.844	1.328	–0.677	1.631
Woman	–1.435	0.318*	–0.945	0.506+
High blood	2.687	0.326*	2.843	0.536*
Diabetes	4.103	0.463*	2.832	0.797*
Cancer	3.757	0.594*	3.421	0.929*
Lung	6.400	0.539*	7.522	0.892*
Heart	7.679	0.462*	8.496	0.945*
Emotional	5.995	0.463*	5.597	0.803*
Oft pain	11.571	0.447*	11.474	0.618*
Interactions with dummy NL				
Constant	–0.955	1.745	–3.064	2.031#
Education med.	2.011	0.883*	2.867	1.025*
Education high	1.937	0.978*	3.613	1.183*
Age 15–44	14.980	6.369*	12.755	8.431#
Age 45–54	3.736	1.716*	2.462	1.960
Age 55–64	1.761	1.734	1.466	2.006
Woman	2.387	0.756*	1.544	0.874+
High blood	–1.729	0.878*	–2.230	1.001*
Diabetes	1.503	1.613	1.418	1.872
Cancer	–1.248	1.521	–0.484	1.742
Lung	0.425	1.354	–1.408	1.621
Heart	1.104	1.287	0.421	1.562
Emotional	2.000	1.027+	1.485	1.240
Oft pain	3.920	0.860*	4.029	0.981*

Panel B Threshold parameters

	γ^1	s.e.	γ^2	s.e.	γ^3	s.e.	γ^4	s.e.
Constant	0.000	0.000	2.017	0.149*	1.988	0.138*	2.101	0.115*
Education med.	–0.932	0.572#	0.044	0.091	0.022	0.090	–0.022	0.078
Education high	–1.149	0.755#	0.054	0.116	0.084	0.112	–0.026	0.097
Age 15–44	1.113	0.814#	0.147	0.134	–0.115	0.144	–0.153	0.130
Age 45–54	1.004	0.710#	0.051	0.118	–0.117	0.115	0.066	0.092
Age 55–64	–0.004	0.738	0.108	0.120	–0.110	0.126	0.035	0.091
Woman	0.602	0.469#	–0.065	0.074	–0.123	0.077#	0.028	0.064
High blood	0.402	0.500	–0.155	0.083+	0.118	0.090#	–0.050	0.073
Diabetes	–1.257	0.748+	–0.016	0.121	0.127	0.124	–0.028	0.109
Cancer	–0.489	0.871	0.082	0.125	–0.033	0.134	–0.121	0.111
Lung	1.528	0.832+	–0.286	0.174+	0.047	0.163	–0.102	0.132

(*continued*)

Table 9.16 (continued)

	γ^1	s.e.	γ^2	s.e.	γ^3	s.e.	γ^4	s.e.
Heart	0.673	1.058	0.071	0.195	–0.351	0.224#	0.123	0.144
Emotional	–0.409	0.706	–0.005	0.117	–0.075	0.139	0.007	0.087
Oft pain	–0.267	0.492	0.079	0.078	0.002	0.082	0.036	0.069
Interactions with dummy NL								
Constant	–2.849	0.886*	0.376	0.147*	–0.062	0.136	0.118	0.113
Education med.	1.016	0.605+	–0.082	0.094	0.036	0.095	0.046	0.082
Education high	1.789	0.781*	–0.072	0.118	–0.043	0.115	0.096	0.100
Age 15–44	–1.830	0.856*	–0.173	0.138	0.084	0.149	0.051	0.134
Age 45–54	–1.039	0.758#	–0.057	0.122	0.062	0.121	–0.263	0.099*
Age 55–64	0.105	0.788	–0.175	0.125#	0.152	0.132	–0.142	0.099#
Woman	–1.050	0.498*	0.095	0.076	0.134	0.081+	–0.012	0.067
High blood	–1.012	0.545+	0.223	0.086*	–0.094	0.094	0.044	0.077
Diabetes	–0.641	0.882	0.109	0.131	–0.107	0.139	0.054	0.124
Cancer	0.986	0.961	–0.142	0.136	0.090	0.149	0.222	0.122+
Lung	–2.422	0.930*	0.309	0.182+	0.003	0.172	0.117	0.140
Heart	–0.421	1.107	–0.090	0.199	0.308	0.229#	–0.202	0.151#
Emotional	–0.669	0.757	0.013	0.122	0.101	0.145	0.037	0.093
Oft pain	0.338	0.528	–0.092	0.081	–0.050	0.087	–0.093	0.074

Panel C Vignette equation

	θ	s.e.
Dummy vig 1	0.800	0.841
Dummy vig 2	5.104	0.863*
Dummy vig 3	16.825	1.098*
Dummy vig 4	16.816	1.097*
Dummy vig 5	14.982	1.052*
V woman	–0.265	0.078*
Sig vig	6.449	0.270*

Panel D Two-point and Five-point scales

	Coefficient	s.e.
λ	0.788	0.046*
σ_{u2}	4.317	0.776*
σ_{u5}	7.213	0.532*

Notes: * denotes $p<0.01$. + denotes p, 00.05. # denotes p, 0.10.

[a] Normalization: $\sigma_r^2 = 10$.

revealed by the estimates for the first threshold equation (γ^1) in panel B; the other threshold parameters appear not to play a large role here.[7] The complete model corrects for this. In the Netherlands, the correlation between

7. A model in which all thresholds shift with respondent characteristics in a parallel manner is statistically rejected against the model presented here, but gives very similar corrections in the work disability equation.

education level and work disability is much weaker, both before and after correcting for DIF.

Age is insignificant in the complete model. Of course, this is related to the fact that health conditions are controlled for directly. The large coefficients on the youngest age group are somewhat misleading since this group is quite small in the U.S. data. The age group forty-five to fifty-four in the United States uses higher thresholds than the fifty-five and over age groups. This is similar to the finding of Salomon, Tandon, and Murray (2004) for mobility (as a domain of general health, not work related), who explains it from expectations: older respondents may more often expect to have some work disability and adjust their scales accordingly. In the model that does not correct for DIF, this would lead to the conclusion that this age group has significantly lower work disability. The role of gender is also smaller in the model that controls for DIF than in the model without DIF.

Health condition dummies are answers to questions of the form, "has the doctor ever told you that . . . ," except for pain, which is self-reported (e.g., "do you often suffer from pain?"). The same variables were used in section 9.3. They are included as exogenous background variables; we assume that these health conditions do not suffer from reporting errors or other measurement errors. Different health conditions have very different effects on work disability, as in the binary probits in the previous section. This does not change much after correcting for response scale differences.

In section 9.3, we found that the effect of pain on reported work disability is much larger in the Netherlands than in the United States. The results in table 9.16 confirm this result. In the United States, pain has a larger effect on work disability than any other health condition. The significantly positive interaction with the dummy for the Netherlands indicates that the effect is even stronger in the Netherlands. Correcting for DIF hardly changes the effect of pain in either the United States or the Netherlands. Thus, differences in response scales for reporting work disability cannot explain why the effect of pain on reported work disability is so much larger in the Netherlands than in the United States.

Panel C contains the estimates for the vignette equations. The dummies for the five vignettes are in line with the idea that vignettes 3, 4, and 5 describe more serious health problems than vignettes 1 and 2. There appears to be a systematic difference between evaluating male and female vignette persons (the parameter on the dummy female in 0). For a given vignette description, a male vignette person is seen as more work disabled than a female vignette person, by both male and female respondents.[8] The estimated standard deviation of the vignette evaluations is much smaller than that of the self-reports. This is in line with the fact that everyone gets the

8. We included an interaction term of respondent gender and gender of the vignette person, but this was insignificant.

same vignette descriptions (apart from the name of the person described, determining the gender). In the self-reports, heterogeneity in respondents' own work disability not explained by gender, education, or age, leads to the much larger variance of the unsystematic part.

Finally, panel D presents the auxiliary parameters related to the transformation between the two-point and the five-point scale. The cut off point for the two-point scale is a weighted mean of the first and second threshold in the five-point scale, with an estimated weight for the first threshold of 0.79. Both idiosyncratic errors in the vignette reports play a role, and are of similar order of magnitude as the unobserved heterogeneity term in true latent work disability, which is common in both reports and has variance 10, by means of normalization.

Table 9.17 compares predictions of work disability for the age group forty-five to sixty-four on the two-point scale of the models with and without DIF (the same two models presented in the first panel of table 9.16). The model without DIF predicts work disability rates of 34.8 percent in the Netherlands and 20.6 percent for the United States, close to the observed work disability rates on the two-point scale for this age group. For the model with DIF, the estimated thresholds for the United States are used. For the U.S. sample, this again closely reproduces the observed work disability rate. This is due to the way the prediction is computed: there is no correction for within U.S. DIF, only for cross-country DIF. For the Netherlands, however, the result is quite different. For every Dutch respondent, the work disability probability is computed as if this respondent would use the threshold of a U.S. respondent with the same characteristics (age, education level, gender, health conditions). The results show that, if the Dutch

Table 9.17 Predicted work disability and health conditions

	Model without DIF		Model with DIF	
	NL	U.S.	NL	U.S.
Total work disability	34.81	20.64	27.64	20.64
Work disability explained by				
Hypertension	0.61	2.09	0.36	2.20
Diabetes	0.73	0.94	0.52	0.66
Cancers	0.28	0.46	0.31	0.42
Lung diseases	0.99	1.13	0.99	1.31
Heart diseases	1.97	2.36	1.99	2.58
Emotional diseases	2.70	1.75	2.39	1.63
Pain	15.21	7.63	14.55	7.56
All health conditions	22.49	16.36	21.12	16.36

Notes: Age group forty-five to sixty-four, CentERpanel and HRS; Weighted using respondent weights. First row: total work disability. Other rows: Reduction in total work disability if dummy for given health condition (or dummies for all health conditions) is always zero. In the model with DIF, work disability is predicted using U.S. response scales.

would use the American thresholds, the self-reported work disability rate in the Netherlands would be reduced to 27.6 percent, a difference of about 7.4 percentage points compared to the 34.8 percent in the model without DIF. Thus, correcting for cross-country DIF reduces the gap between the United States and the Netherlands from 14.2 percentage points to 7.0 percentage points, a reduction of about 50 percent.

The other rows in table 9.17 predict how much each health condition contributes to explaining work disability according to both models, again using U.S. response scales for the model with DIF. Work disability is recomputed after setting the dummy for the given health condition equal to zero, and the reduction in work disability compared to the first row is reported. The differences between the two models are small. Pain remains the dominating factor in both countries, and is much more important in the Netherlands than in the United States. Thus we find that there is a considerable difference in response scales between Dutch and U.S. respondents explaining a large part of the observed difference in the work disability rate, but the difference is not related to whether respondents suffer from a health condition or not. All health conditions together explain most of reported work disability according to both models. They explain more in the Netherlands than in the United States, again due to the effect of pain.

Table 9.18 gives the prevalence rates of the health conditions in the age group forty-five to sixty-four and the average marginal effect of each health condition on the probability of work disability. As in table 9.17, the estimated U.S. response scales are used for both the Dutch and the American respondents. Table 9.17 decomposes the contributions to work disability in two components: prevalence and the marginal effect. There are some differences between the models that do and do not correct for DIF across

Table 9.18 **Prevalence and marginal effects**

	Prevalence (in %)		Average marginal effect (%-points)			
			Model without DIF		Model with DIF	
	NL	U.S.	NL	U.S.	NL	U.S.
Hypertension	25.38	36.04	2.41	5.80	1.43	6.10
Diabetes	4.64	9.16	15.69	10.24	11.27	7.18
Cancer	4.53	5.25	6.20	8.73	6.85	7.98
Lung disease	6.35	6.84	15.52	16.55	15.67	19.20
Heart disease	8.42	11.69	23.40	20.21	23.67	22.07
Emotional disorder	12.81	11.14	21.10	15.69	18.63	14.66
Pain	32.09	24.07	47.41	31.71	45.35	31.43

Notes: Age group forty-five to sixty-four, CentERpanel and HRS; Weighted using respondent weights. Prevalence: fraction of the sample with the given health condition. Average marginal effect taken over all observations with given health condition.

countries, but the qualitative conclusions remain the same. Pain has both the largest prevalence rate and the largest marginal effect in both countries, explaining why it has by far the strongest contribution on work disability. In the Netherlands, both prevalence and marginal effect are substantially larger than in the United States, explaining why the contribution of pain to explaining work disability is larger in the Netherlands than in the United States.

9.6 Conclusions

Workers in different industrial western countries report very different rates of work disability. The diversity in reported work disability stands in sharp contrast to the believed relative similarity in their observed health outcomes. This contradiction continues to be seen as a major unresolved puzzle.

In this chapter, we investigated the role of pain as a factor leading to work disability in three countries—the Netherlands, the United Kingdom, and the United States. In all three countries, pain is by far the most important factor leading to reports of work disability. We also find, however, that respondents in these three countries who appear to be suffering from similar degrees of pain respond very differently to questions on work disability. These differences do not appear to be related to differential use of painkillers to alleviate the effects of pain or differential degrees of work accommodation available in the three countries.

Using a new methodology of vignettes which were implemented in Internet surveys in the United States and the Netherlands, our analysis claims that a significant part of the observed difference in reported work disability between the two countries is explained by the fact that residents of the two countries use different response scales in answering the standard questions on whether they have a work disability. Essentially for the same level of actual work disability, Dutch respondents have a lower response threshold in claiming disability than American respondents do. An important follow-up question is what causes these differences in thresholds across countries.

One possibility is that more people in the Netherlands know people who are on work disability programs than is the case for residents of the United States and that this familiarity makes people less tough on what it takes to constitute a work disability. In a recent paper, Van Soest et al. (2007) do find that reference group effects (the fraction of people one knows on work disability) are significant, and contribute substantially to an explanation of why self-reported work disability in the Netherlands is much higher than in the United States. This implies an important role for social interactions and norms on the perception of work limitations.

References

Bound, J., and R. Burkhauser. 1999. Economic analysis of transfer programs targeted on people with disabilities. In *Handbook of labor economics, Vol. 3C*, ed. O. Ashenfelter and D. Card, 3417–3528. Amsterdam: Elsevier.

Browning, M., T. Crossley, and G. Weber. 2003. Asking consumption questions in general purpose surveys. *Economic Journal* 113:F540–67.

Burkhauser, R., M. Daly, A. Houtenville, and N. Nargis. 2002. Self-reported work limitation data—what they can and cannot tell us. *Demography* 39 (3): 541–555.

Haveman, R., V. Halberstadt, and R. Burkhauser. 1984. *Public Policy Toward Disabled Workers.* Ithaca and London: Cornell University Press.

Kapteyn, A., J. P. Smith, and A. van Soest. 2007. Self-reported work disability in the US and the Netherlands. *American Economic Review* 97 (1): 461–73, March 2007.

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): 191–207.

National Center for Health Statistics. Data File Documentation, National Health Interview Survey, 2002 (machine readable data file and documentation). Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention.

Salomon, J., A. Tandon, and C. Murray. 2004. Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. *British Medical Journal* 328 (7434): 258–60.

Van Soest, A., A. Andreyeva, A. Kapteyn, and Smith, J. P. 2007. Self reported disability and reference groups. RAND Working Paper no. WR-409-1.

Ware, J. J., and C. D. Sherbourne. 1992. The MOS 36-item short form health survey (SF-36) I: Conceptual framework and item selection. *Medical Care* 30:473–83.