AW
7/25/06
3:10 PM

Self Reported Disability and Reference Groups

Arthur van Soest, RAND and Tilburg University
Arie Kapteyn, RAND Tatiana Andreyeva, RAND
Jim Smith, RAND

July 20, 2006

Abstract

Different people may have different thresholds for the extent to which health problems limit the kind and amount of work they can do. The thresholds may be related to the kind of work one does currently, demographics, work history. In this paper we look at a different determinant of self-reported disability, namely social influence. Using data on self-reports of work disability, work disability vignettes, and self-reports on disability in one's reference group we estimate a model describing the influence of work disability prevalence in one's reference group on the subjective scale used to report own work disability. It appears that these reference group effects exert a significant influence, thereby suggesting a significant role for social influences (and perhaps social norms) on the perception of work limitations.

1 Introduction

Both prevalence of Disability Insurance (DI) benefit receipt and self reported work disability vary substantially across countries, as documented for instance in Haveman and Wolfe (2000) , Bound and Burkhauser (1999). Banks et al. (2005) and Kapteyn, Smith, and Van Soest (2006) have explored to which extent differences in self reported work disability may be ascribed to differences in reporting styles across countries. They have exploited a vignette methodology developed by King et al. (2004) and applied by Salomon, Tandon, and Murray (2004).

Kapteyn, Smith, and Van Soest (2006) report that in the age bracket 51-64 self reported work disability in The Netherlands is about 58% higher than in the United States (35.8% of the respondents in the age group 51-64 in The Netherlands report

some work disability against 22.7% in the U.S.). They find that more than half of the observed difference can be explained by differences in response scales. That is, if the Dutch respondents are assigned the response scales of U.S. respondents then the Dutch prevalence of self-reported work disability falls from 35.8% to 28.3%.

This result generates a number of obvious follow-up questions, including (1) what explains the remaining difference in self-reported work disability and (2) where do the differences in reporting scales come from. In this paper we do not attempt to answer the first question, but instead put forward and test a hypothesis regarding the second question. The basic idea underlying our hypothesis is straightforward. For various reasons, DI benefit receipt is much higher in The Netherlands than in the U.S. Bound and Burkhauser (1999) report that in 1995, the number of DI recipients per 1000 workers in the age group 45-59 was 103 in the U.S., compared to 271 in The Netherlands. Factors contributing to this difference are higher benefits and less strict eligibility rules in The Netherlands compared to the U.S. See for instance Aarts, Burkhauser and De Jong (1996). [1]

It is *not* very likely that differences in health explain the difference in self-reported work disability. An analysis of a broad set of health conditions by Banks et al. (2005) suggests that the Dutch population is actually healthier than the U.S. population. It seems likely therefore that at a given health level a Dutch individual has a higher likelihood of receiving DI benefits than his or her U.S. counterpart. This in itself may induce a difference in social norms, whereby a higher prevalence of DI benefit receipt leads an individual to define a given health condition more easily as work limiting, essentially because he or she is more likely to know someone else with a similar condition receiving DI benefits.

We formalize this notion by introducing the concept of prevalence of DI benefit receipt in one's *reference group.* A reference group is defined as one's circle of friends and acquaintances. In a Dutch survey we ask respondents directly how many people among their friends and acquaintances receive DI benefits. In this paper, we develop a model that jointly explains the categorical answer to this question and self-reported work disability. The main feature of the model is the notion that response scales for reporting no, mild, or severe work disability, can be affected by the "peer group effect," i.e., by the number of people in the reference group receiving disability benefits. To identify the determinants of response scales, we exploit anchoring vignettes as in Kapteyn, Smith, van Van Soest (2004).

[1] In 2004, DI recipients in The Netherlands made up 13% of the labor force (Source: Statistics Netherlands http://statline.cbs.nl/StatWeb.)

In the US the percentage of DI-recipients was 4.8% of the civilian labor force (Source: US Bureau of Labor Statistics ftp://ftp.bls.gov/pub/news.release/History/empsit.01072005.news)

In the next section we briefly describe the micro-data used in our analysis. Section 3 presents the model, which essentially consists of three equations. The first equation explains the answers of respondents to the question about DI benefit receipt in their reference group. The second equation models self-reported work disability. The third equation (or rather set of equations) explains how individual response scales to questions on work disability (or anchoring vignettes) are affected by the prevalence of DI benefit receipt in the reference group. Throughout we control for a large number of other variables, such as socio-demographic characteristics and health conditions. Section 4 presents the results of model estimation. We find that DI benefit receipt in one's reference group has a significant effect on response scales in the expected direction. To gauge the size of this effect, we graph the relation between DI benefit receipt in the reference group against self-reported work disability. It turns out that to explain the complete difference in response scales between the U.S. and The Netherlands, the percentage of respondents in The Netherlands reporting to know at least some DI benefit recipients has to fall by about half. This is an order of magnitude that seems reasonable given the substantial difference in the number of Dutch and U.S. people on DI benefits, although it remains to be confirmed using similar data for the U.S. Section 5 concludes.

2 The Data

We use information obtained from the Dutch CentERpanel. The CentERpanel is an Internet panel of about 2,250 households who have agreed to respond to a survey every weekend. The sample is not restricted to households with their own Internet access. Respondents are recruited by telephone. If they agree to participate and do not already have Internet access, they are provided with Internet access (and if necessary, a set-top box). Thus, the CentERpanel is representative of the Dutch adult population except the institutionalised. Sample weights based upon data from Statistics Netherlands are used to correct for unit nonresponse. The sample that we use to estimate our model consists of about 2,000 respondents who participated in several interviews with questions on work disability in 2003.

From multiple waves of the data that have been collected in the past, the CentERpanel has a rich set of variables on background and demographic characteristics of the respondent and household, their income and labor market status, and several salient dimensions of health. In August 2003, we collected work disability self-reports and vignette evaluations (described below) in the CentERpanel. The Internet infrastructure makes the CentERpanel an extremely valuable tool to conduct experiments, with possibilities for randomization of content, wording, question and response or-

der, and regular revisions of the design. Production lags are very short, with about one month between module design and data delivery. For example, based upon our initial analysis, we fielded a second wave in October with slightly different wordings of the vignette questions. A third wave of experiments was administered in December 2003. In the October wave we also included a number of questions about reference groups, which will be exploited in the current paper. For the current analysis we will use the data from the 2003 October wave. Appendix A lists the vignette questions. All vignettes are presented with either a female or a male name.[2]

For each of the vignettes the respondent is asked the following question: "Does ... have a health problem that limits the amount or type of work he/she can do?" with a five point response scale: not at all; yes, mildly limited; yes moderately limited; yes, severely limited; yes, extremely limited/cannot work. Table 1 presents the response frequencies for each of the 15 vignette questions given in Appendix A.

[2]Female or male names are assigned randomly. In Appendix A we only show one of the two names per vignette. A further random variation in question wording is that respondents are either told that the persons in the vignettes have a specific job (e.g. bus driver, bank teller) or are told the vignette person "has the same job and a similar career as you have had". For the current analysis we ignore this variation and make the assumption that this will not have an appreciable effect on the outcomes of the analysis.

Table 1. Frequencies for Vignette Answers (CentERpanel, October 2003)

Affect vignettes	Affect 1	Affect 2	Affect 3	Affect 4	Affect 5
Not at all limited	41.2	96.2	11.1	18.7	2.2
Somewhat limited	49.7	2.8	44.3	44.8	8.4
Moderately limited	7.2	0.6	31.2	26.0	18.6
Severely limited	1.4	0.5	12.2	8.9	40.4
Extremely limited/cannot work	0.5	0.0	1.3	1.6	30.4

Pain vignettes	Pain 1	Pain 2	Pain 3	Pain 4	Pain 5
Not at all limited	22.5	8.2	0.6	0.3	0.8
Somewhat limited	61.8	47.1	6.6	6.2	12.9
Moderately limited	13.4	34.1	25.7	29.4	31.3
Severely limited	1.9	9.2	49.5	43.2	39.2
Extremely limited/cannot work	0.4	1.4	17.6	20.9	15.9

CVD vignettes	CVD 1	CVD 2	CVD 3	CVD 4	CVD 5
Not at all limited	91.2	10.6	1.8	20.7	6.7
Somewhat limited	7.8	46.2	18.2	44.9	34.1
Moderately limited	0.9	29.2	32.6	25.0	30.3
Severely limited	0.1	11.8	33.6	8.8	20.7
Extremely limited/cannot work	0.0	2.3	13.9	0.6	8.3

Notes: Data are weighted. Complete sample N=1980.
See Appendix 1 for the wordings of the vignette questions.

The distribution of answers corresponds quite well with the severity of the conditions described in the different vignettes. Banks et al. (2005) present descriptive tables of how respondents rank the vignettes according to severity.[3] There appears to be strong agreement among respondents regarding the ordering of vignettes in terms of severity.

Table 2 presents the distribution of the answers to the question on work limitations by age group. These are the answers to the question: "Do you have an impairment or health problem that limits you in the amount or kind of work you

[3]These were the vignettes from the August wave.

can do?". The question allows respondents to reply on a 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, (5) Yes, I am very severely limited-I am unable to work. These response categories are identical to the ones used to gauge the severity of the vignette work limitations. The table implies that about 37% of the Dutch population has at least a mild work limitation. About 14% have a work limiting health problem or impairment that they gauge as moderately limiting or worse. Work related health seems to deteriorate with age (although we cannot exclude cohort effects either). The most interesting groups here are probably people in the age groups 45-54 and 55-64. For them, the prevalence of work limiting health problems are large, and this might be an important reason not to participate in the labor market. For the 65-plus, work limiting health problems are still more prevalent, but these people are almost always retired anyhow in the Netherlands.

Table 2. Distribution of Self-Reported Work Disability by Age, %

	Age Group						
	15-24	25-34	35-44	45-54	55-64	65+	Total
Not at all limited	86.8	74.1	69.2	55.9	52.8	48.4	63.1
Somewhat limited	5.4	20.7	17.5	24.2	28.5	34.3	22.8
Moderately limited	5.8	3.2	5.8	7.0	10.5	10.9	7.1
Severely limited	2.0	0	2.1	2.9	1.8	3.7	2.2
Extremely limited/cannot work	0	1.8	5.4	9.9	6.3	2.8	4.8
Number of observations	68	362	438	460	336	316	1980

Notes: Data are weighted. Complete sample N=1980.

Appendix B presents some of the questions about reference groups asked in the October wave. We only report the questions that are also used in the empirical analysis.

Our operationalization of a reference group is the circle of acquaintances mentioned in these questions. The first two reference group questions provide information on the modal age and modal education level of the respondent's reference group. In the analysis we will combine the age categories into a smaller number of broader age brackets; similarly for the education categories. Table 3 presents descriptive statistics for the independent variables used in this study, including the responses to the

first two reference group questions listed in Appendix B. For example, 27 percent of all respondents report that most of the people in their reference group are in the age group 36-45. About 48 percent say that most of their reference group have medium education level (while 39 percent of the respondents has that level).

Table 3. Sample Statistics for Independent Variables

	Mean/percentage
Stroke	1.3
Cancer	3.8
Lung disease	6.0
Heart disease	7.1
High blood pressure	19.2
Diabetes	4.8
Emotional problems	11.0
Arthritis	10.4
Problems with vision	3.8
Often pain	25.4
Age in years	47.6
Low education level	39.1
Medium education level	38.7
High education level	22.1
Female	49.9
Northern provinces*	14.3
Eastern provinces*	21.6
Western provinces*	38.7
Southern provinces*	25.5
Age in reference group <25	8.7
Age in reference group 25-35	20.2
Age in reference group 36-45	27.0
Age in reference group 46-55	19.7
Age in reference group 56-65	14.7
Age in reference group 66+	9.8
Low education level in the reference group	24.9
Medium education level in the reference group	47.9
High education level in the reference group	27.2

Notes: Data are weighted. Estimation sample N=1764.
All variables other than "Age in years" are dummies. The table gives the percentage of observations for which the dummy has value 1.

*Northern provinces are Groningen, Friesland & Drenthe
*Eastern provinces are Overijssel, Flevoland & Gelderland
*Western provinces are Utrecht, Noord-Holland & Zuid-Holland
*Southern provinces are Zeeland, Noord-Brabant & Limburg

The other reference group questions refer to the number of acquaintances receiving disability benefits, separately for men and women. These are crucial variables for our analysis, measuring disability in the reference group. For men, we will use the number of male acquaintances on disability benefits; for women, we will only consider the female acquaintances. The distribution of these answers by gender and age group is presented in Table 4. Here and in the rest of the paper we combine the categories of prevalence of DI-receipt in the reference group to three: "Nobody", "Very Few", "A Few/Many". Young people typically know very few people on disability benefits. The number of reference group members on disability benefits is highest for 55-64 year old respondents, who also most commonly receive disability benefits themselves. People older than 65 may often have a work disability (see Table 2) but hardly ever receive disability benefits (they receive state and employer provided pensions instead). The number of women on disability benefits in women's reference groups is typically smaller than the number of men on disability benefits in men's reference groups, particularly at older ages. This may be because women in older cohorts often stopped working at an early age (to raise children) and would never qualify for disability benefits after that.

Table 4. Distribution of Disability in the Reference Group by Age, %

	Men, Age Group						
	15-24	25-34	35-44	45-54	55-64	65+	Total
None	82.9	65.6	52.5	55.1	39.4	53.8	56.7
Very few	17.1	31.5	41.5	36.6	44.1	34.7	35.5
A few/many	0	2.9	5.9	8.4	16.5	11.4	7.8
No of observations	29	174	221	248	196	199	1067

	Women, Age Group						
	15-24	25-34	35-44	45-54	55-64	65+	Total
None	76.4	67.8	60.7	62.6	58.9	55.2	62.6
Very few	23.6	29.0	35.7	30.4	32.9	38.2	32.4
A few/many	0	3.2	3.6	7.1	8.2	6.5	5.0
No of observations	39	188	217	212	140	117	913

Notes: Data are weighted. Complete sample N=1980.

It seems plausible that these reference group variables are endogenous to the respondent's work disability – respondents who have a work disability will often not work and not only receive disability benefits, but also more easily get acquainted with other people on disability benefits. This is why we will treat the number of acquaintances on disability benefits as a dependent variable in our model. Since our primary interest is in work limitations, Table 5 shows cross tabs of self reported work limitations and self reported incidence of DI-receipt in one's reference group. We combine categories for self reported work disability to three: "Not Limited", "Mildly Limited", "Moderately Limited/Severely Limited/Extremely Limited". The table clearly illustrates a positive relation between self reported work limitations and the number of people in one's reference group drawing disability benefits. Of course there are several competing explanations for this positive association. First of all, there may be a true causal effect of the prevalence of DI-receipt in one's reference group on the tendency to report work limitations. Secondly, as discussed above, it is possible that respondents with work limitations are more likely to associate with others who have a work disability (e.g. because of the existence of networks of people with work disabilities). Thirdly, there may be other factors that both increase the likelihood that a respondent has a work limitation and that she or he knows others with work limitations. One such factor is age. Fourth, responses may be biased for several reasons. Respondents may, for instance, exaggerate the number of friends or acquaintances on DI to "justify" their own report of a work limitation (cf. Bound, 1991). These explanations are not mutually exclusive and undoubtedly there are more. In the next section we present a model that aims at isolating the importance of the first of the explanations given here. In the discussion of the results we will return to the competing explanations.

Table 5. Self-Reported Work Disability and Reference Group Disability

		Disability in the reference group, %			
		None	Very few	A few/many	***Total***
Self-reported work disability, %	Not limited	39.6	20.1	2.8	62.5
	Mildly limited	12.1	8.9	2.3	23.3
	Moderately, severely and extremely imited	6.4	5.5	2.3	14.2
	Total	58.1	34.6	7.3	100.0

Notes: Data are weighted. Estimation sample N=1764.

3 A Model with Reference Groups

We propose a fairly simple econometric model to explain the reported number of people on disability in the reference group R, self-reported work disability Y, and disability of vignette persons $Y^1, \ldots, Y^L$ ($L = 15$ in our case).

Disability in the reference group

We only consider disability in the respondent's reference group of the respondent's own sex. As mentioned, we combine the outcomes "few" and "many" because of the small number of observations with the latter outcome. Thus we obtain an ordered response variable with three possible outcomes, let's say $j = 1$ ("none"), $j = 2$ ("very few") and $j = 3$ ("a few" or "many"). This will be modelled with an ordered probit equation

$$R^* = X^R\beta^R + \omega^R, \quad \omega^R \sim N(0, \sigma_\omega^2) \tag{1}$$

$$R = j \text{ if } \phi_{j-1} < R^* \leq \phi_j, \ j = 1, 2, 3. \tag{2}$$

For notational convenience, we define $\phi_0 = -\infty$ and $\phi_3 = \infty$. Below we will further specify the parameters ϕ_1 and ϕ_2. The vector X^R of respondent characteristics driving reference group disability is assumed to be independent of all the errors in

the model. Equation (1) has a "reduced form" nature in the sense that we do not explicitly model how work disability or labor force status affects disability in the reference group. The exogenous determinants of labor force status and disability are included among the regressors X^R to account for this.

Self-reports

Self-reports and vignette evaluations are modelled in the same way as in the King et al.(2004) model. For self-reported work disability Y on a 3-point scale (recall that we combined the three most serious categories "moderate," "severe," and "extreme," to one) we have:

$$Y^* = X\beta + \epsilon \tag{3}$$

$$Y = j \text{ if } \tau_{j-1} < Y^* \leq \tau_j,\ j = 1, 2, 3. \tag{4}$$

Also here, for notational convenience, we define $\tau_0 = -\infty$ and $\tau_3 = \infty$. The remaining thresholds τ_1 and τ_2 are modeled as functions of observable and unobservable respondent characteristics; see below. The error term ϵ is assumed to be standard normally distributed. We allow for correlation between ϵ and ω^R since there will be common factors in both and to allow for the role of actual labor force status (which is not included explicitly in the model but "substituted out"): work disability drives labor force status, and labor force status drives the reference group. We will assume that ϵ and ω^R follow a bivariate normal distribution with correlation coefficient ρ. The vector X of respondent characteristics driving work disability is assumed to be independent of all the errors in the model.

Vignettes

The evaluations Y^l of vignettes l, $l = 1, \ldots, L(= 15)$, are given by

$$Y^{l*} = \theta^l + \delta F^l + \epsilon^l \tag{5}$$

$$Y^l = j \text{ if } \tau_{j-1} < Y^{l*} \leq \tau_j,\ j = 1, 2, 3. \tag{6}$$

Here F^l is a dummy variable indicating whether the person described in the vignette is female ($F^l = 1$) or not ($F^l = 0$). This specification follows earlier work by Kapteyn, Smith, Van Soest (2006) who find that respondents (both male and female) tend to be "harsher" on vignette persons with a female name than with a male name. That is, δ is found to be negative. We assume that all ϵ^l are independent of each other and of the other error terms, and follow a normal distribution with mean zero and variance σ_v^2. Thus the ϵ^l are interpreted as idiosyncratic noise driving vignette

evaluations; they reflect arbitrariness in each separate evaluation. If respondents have a persistent tendency to give low or high evaluations, this will be captured by an unobserved heterogeneity term in the response scales, see below.

Response scale thresholds

The thresholds τ_1 and τ_2 are modeled as follows:

$$\tau_1 = V\gamma_1 + \gamma_1^R R^* + \xi \tag{7}$$

$$\tau_2 = \tau_1 + e^{V\gamma_2 + \gamma_2^R R^*} \tag{8}$$

We have included a vector V of respondent characteristics (independent of all error terms) to allow for a rather general way in which response scales vary with individual characteristics. The distance between the two thresholds is also allowed to depend on these characteristics. The exponential forces it to be positive (as in King et al., 2004). The parameters of interest are γ_1^R and γ_2^R. In particular, γ_1^R is expected to be negative: people who know many people on disability benefits will think of work disability as something common and will more often evaluate people (including themselves) as work disabled, thus using lower thresholds.[4]

The term ξ reflects unobserved heterogeneity in thresholds. For computational convenience, we assume that that there is no unobserved heterogeneity in the distance between the two thresholds. ξ is assumed to follow a normal distribution with variance σ_ξ^2, independent of V, X and X^R, and of the other error terms in the model, ω^R, ϵ, ϵ^1, ..., and ϵ^L. We allow for correlation between the thresholds τ_j and the thresholds in the reference group equation ϕ_j by specifying: $\phi_1 = \phi_{01} + \mu\xi$, $\phi_2 = \phi_{02} + \mu\xi$. The reason for this specification is that, for example, people who tend to answer "many" to the reference group question may also overrate their own work disabilities. We normalize $\phi_{01} = 0$, and define $\phi_{00} = -\infty$, $\phi_{03} = \infty$.

Define $u^R = \omega^R - \mu\xi$. For normalization we set $\text{Var}(u^R) = 1$. We can then write (2) as

$$R = j \text{ if } \phi_{0,j-1} < X^R\beta^R + u^R \leq \phi_{0j},\ j = 1, 2, 3. \tag{9}$$

where ϕ_{02} is the only parameter to be estimated. Note also that the correlation between ε and u^R equals ρ. The assumption that ξ is independent of ϵ implies that people with higher thresholds do not tend to have larger or smaller genuine work disability (on a continuous scale), keeping observed characteristics X and V constant. This assumption helps in identification and avoids the need for exclusion

[4] In the empirical work, we will allow the parameters γ_1^R and γ_2^R to depend on education level. For notational convenience, we do not make this explicit in the notation.

restrictions. The assumption seems quite plausible, although one might argue that lower thresholds point at unobserved characteristics such as pessimistic views that can also genuinely reduce respondents' ability to work.

Likelihood Contributions

Compared to the models in King et al. (2004) and Kapteyn, Smith, and Van Soest (2006), there are two complications: the thresholds now depend on an unobserved variable R^* and upon an unobserved heterogeneity term ξ. Replacing R^* using (1) and exploiting (7) and (8) gives:

$$\tau_1 = V\gamma_1 + \gamma_1^R X^R \beta^R + \xi + \gamma_1^R(u^R + \mu\xi), \tag{10}$$

$$\tau_2 = \tau_1 + e^{V\gamma_2 + \gamma_2^R X^R \beta^R + \gamma_2^R (u^R} + \mu\xi) \tag{11}$$

(3) and (4) imply

$$Y = j \text{ if } \tau_{j-1} - X\beta < \epsilon < \tau_j - X\beta \tag{12}$$

Similarly, for the vignette evaluations we get:

$$Y^l = j \text{ if } \tau_{j-1} - \theta^l - \delta F^l < \epsilon^l < \tau_j - \theta^l - \delta F^l \tag{13}$$

The probability of observing a certain reference group category follows from (9):

$$R = j \text{ if } \phi_{0,j-1} - X^R\beta^R < u^R < \phi_{0j} - X^R\beta^R \tag{14}$$

Let the reported reference group variable be r, the observed work disability self-report y, and the observed vignette evaluations $y^1, \ldots, y^L$. Then the likelihood contribution of a given respondent can be written as a two-dimensional integral over the values of u^R that result in $R = r$ and all possible values of ξ:

$$\int_{-\infty}^{\infty} \int_{\phi_{0,j-1} - X^R\beta^R}^{\phi_{0j} - X^R\beta^R} P(Y = y|u^R, \xi) \prod_{l=1}^{L} P(Y^l = y^l|u^R, \xi) f(u^R|\xi) \mathrm{d}u^R \frac{1}{\sigma_\xi} \phi(\frac{\xi}{\sigma_\xi}) \mathrm{d}\xi \tag{15}$$

where ϕ is the standard normal density and f is the conditional density of u^R given ξ, which is univariate normal. Of course, the crucial point here is that, conditional on u^R and η, all vignette evaluations and the self-report are mutually independent, allowing for the factorization in (15). The conditional probabilities in (15) follow from (12) and (13), together with the normality assumptions on the error terms, implying that the ϵ^l are independent of ϵ, ξ and u^R but that $\epsilon|(u^R, \xi) \sim N(\rho u^R, 1 - \rho^2)$:

$$
\begin{aligned}
P(Y = y|u^R, \xi) &= \Phi([\tau_y - X\beta - \rho u^R]/\sqrt{[1-\rho^2]}) \\
&- \Phi([\tau_{y-1} - X\beta - \rho u^R]/\sqrt{[1-\rho^2]})
\end{aligned}
$$

$$
\begin{aligned}
P(Y^l = y^l|u^R, \xi) &= \Phi([\tau_{y^l} - \theta^l - \delta F^l]/\sigma_v) \\
&- \Phi([\tau_{y^l-1} - \theta^l - \delta F^l]/\sigma_v)
\end{aligned}
$$

where the $\tau_{...}$ are given by (10) and (11) (and depend on ξ and u^R).

4 Results

The integrals in (15) have been replaced by smooth simulation, by drawing 1600 (40x40) times from the joint distribution of ξ and u^R. Experiments with a substantially larger number of draws did not lead to appreciable differences in the estimation outcomes, from which we infer that the number of draws is large enough to provide an accurate approximation of the integral. Table 6 presents the estimation results for equations (3) and (1). Estimates for the threshold equations (10) and (11) are given in Table C.1 in Appendix C, while the estimates for (5) are given in Table C.2.

Table 6: Estimation Results for Own and Reference Group Disability

	Self-Reported Disability			Reference Group Disability		
	Coef.	s.e.	t-value	Coef.	s.e.	t-value
age	0.281	0.146	1.922	0.267	0.169	1.583
age squared	-0.020	0.014	-1.407	-0.017	0.016	-1.029
medium education	-0.061	0.084	-0.730	0.067	0.076	0.877
higher education	-0.274	0.085	-3.243	-0.041	0.081	-0.509
female	0.020	0.070	0.285	-0.349	0.063	-5.574
stroke	1.307	0.335	3.904	-0.066	0.251	-0.265
cancer	0.335	0.145	2.314	-0.161	0.161	-0.994
lung	0.621	0.131	4.743	0.301	0.134	2.254
heart problems	0.880	0.130	6.756	-0.002	0.119	-0.021
highblood	0.044	0.082	0.535	0.066	0.075	0.876
diabetes	0.403	0.171	2.359	0.168	0.154	1.091
emotional problems	0.626	0.097	6.447	0.257	0.099	2.587
arthritis	0.424	0.117	3.625	0.204	0.109	1.874
vision	0.098	0.177	0.556	0.027	0.165	0.166
often pain	1.237	0.081	15.249	0.248	0.077	3.241
intercept	-1.969	0.375	-5.256	-1.067	0.384	-2.779
reference group age 25-35				0.042	0.116	0.357
reference group age 36-45				0.161	0.139	1.155
reference group age 46-55				0.229	0.159	1.439
reference group age 56-65				0.055	0.164	0.337
reference group age >65				-0.087	0.178	-0.488
Medium education in R.G.				-0.018	0.055	-0.327
High education in R.G.				-0.063	0.066	-0.955
northern provinces				0.000	0.071	0.000
eastern provinces				0.029	0.062	0.474
western provinces				-0.138	0.058	-2.395
ρ	0.044	0.039	1.127			
φ_{02}	1.334	0.051	26.405			

We first briefly discuss the estimates in Tables C.1 and C.2. A negative sign in Table C.1. means that the corresponding variable leads to a lower threshold and hence to a higher probability that a respondent will label a certain condition as work limiting. Although several of the variables are statistically significant, the magnitude

of most of the coefficients is rather modest. Regarding the first threshold, we observe that females are more likely to judge a given condition to be work limiting, whereas people with higher education are less likely to do so. The age function has a top at 59 years. So until age 59 older people are "tougher", i.e. are less likely to call a condition work disabling. [5] For the distance between the first and second threshold (γ_2), results are somewhat different; the age function has a minimum at 51 years of age, while higher education leads to a smaller distance between thresholds. The estimates are difficult to interpret individually, due to the complexity of the model, where the same variables appear in a number of different equations.

Unobserved heterogeneity in thresholds is significant - the estimated standard deviation of ξ is 0.54 and seems to be very accurately determined. To judge its size, it can be compared to the amount of idiosyncratic noise in self-reports and vignette evaluations. The former has standard deviation 1 (by normalization), the latter has standard deviation 0.52 (see Table C.2). Thus unobserved heterogeneity in thresholds explains about 23% of the unsystematic variation in self-reports and about 52% of the unsystematic variation in vignette evaluations.

The dummy coefficients in Table C.2. reflect the average severity of the work limitations described in the vignettes. One can relate the dummy coefficients to the relative frequencies in Table 1. For instance, the dummy coefficient for vignette number 2 is equal to -1.29, indicating that the person described in this vignette has the least work limiting health condition of all vignettes. This is confirmed by Table 1, where vignette 2 (Affect 2) shows the highest percentage of respondents saying that this person is not at all limited in the kind or amount work he/she can do. In contrast, the dummy coefficients for vignettes 8 and 9 are the highest. Table 1 shows that the health conditions described in these vignettes (Pain 3 and Pain 4) are considered by respondents to be the most work limiting of all. The estimate of δ, the coefficient of the dummy for a female vignette name is small and insignificant, in contrast to earlier findings by Kapteyn, Smith, and Van Soest (2006). The estimated idiosyncratic variation in vignette evaluations σ_v (independent across vignettes) is smaller than the unsystematic variation in self-assessments ($\sigma_\epsilon = 1$, by means of normalization). This makes sense since self-reports not only have reporting noise but also genuine unobserved heterogeneity in work related health, while the vignette descriptions control for that.

The equation for self-reported disability in Table 6 shows that self-reported disability goes up with age until age 71; it is lower for higher educated individuals and higher for individuals with serious health conditions, including strokes, heart problems, cancer, diabetes, emotional problems, and lung problems. The reference group

[5] The variable "age" is defined as age divided by 10.

disability equation shows that the reported DI prevalence in the reference group increases with age (the top of the parabola is estimated at 80 years); it shows virtually no relation with own education, and implies that DI receipt in the reference group increases with most health conditions, in line with the argument that people with a health problem will more often be acquainted with other people in poor health. At first sight surprisingly, reported DI prevalence falls once the modal reference goup age exceeds 65. As already discussed in the data section, this may well be an artefact of the Dutch DI system, where anyone over 65 is ineligible for DI benefits but receives income from other sources. Also in line with the raw data (Table 4) is that females are less likely to report DI-benefit recipients in their reference group. Respondents in the western provinces are less likely to know people on disability benefits than respondents in the rest of the country.

The parameters of primary interest are γ_1^R and γ_2^R. Both have been parameterized as a function of education level, see Table 7. For individuals with lower education γ_1^R is estimated at -.43; for individuals with medium education the estimate is -.40, while for the higher educated the estimate is -.33. The negative values imply that respondents with many people on DI in their reference group are more likely to consider a given condition as work limiting. This applies less to the higher than to the lower educated. The estimates for γ_2^R show that the distance between thresholds increases with the number of acquaintances on disability benefits, particularly for the higher education levels.

Table 7: Estimation Results for Reference Group Effects

	γ_1^R			γ_2^R		
	Coef.	s.e.	t-value	Coef.	s.e.	t-value
medium education	0.027	0.026	1.031	-0.060	0.018	-3.284
higher education	0.103	0.027	3.863	-0.037	0.018	-2.028
intercept	-0.434	0.125	-3.472	0.096	0.014	6.791

We note from Table 6 that the parameter ρ, the correlation between the error terms in equations (1) and (3) is small and insignificant. Thus the correlation in the error terms between self-reported disability and DI-benefit receipt in the reference group appears to be negligible. The standard deviation of the individual effect in the response scales (σ_ξ in Table C.1) is estimated at .54, while μ is estimated at -.82. In view of the fact that $u^R = \omega^R - \mu\xi$ and that var(u^R)= 1, we find that var(ω^R)=.80. The sign of μ is counter intuitive, as it implies that if a respondent uses high thresholds for answering about his or her own work limitations, he or she will tend to use low thresholds when asked for DI prevalence in the reference group.

This would be the opposite of justification bias.

Table 8 provides a simple way of checking the fit of the model. It is similar to Table 5, but reports frequencies simulated using the model. Comparing the two tables suggests that the fit of the model is fairly good; judging from the marginals, the model does a good job in replicating reported reference group DI-receipt; it does a slightly worse job in reproducing the distribution of self-reported disability. The biggest deviation between the data and the model predictions occurs in the middle category (mildly limited). According to the data, 23.3% of the respondents classify themselves as mildly limited, whereas the model predicts 19.9% in that category.

Table 8. Model Predictions of Self-Reported Work Disability and Reference Group Disab

		Disability in the reference group, %			
		None	Very few	A few/many	***Total***
Self-reported work disability, %	Not limited	40.7	21.1	3.8	65.6
	Mildly limited	10.2	7.7	2.0	19.9
	Moderately, severely and extremely limited	6.2	6.4	1.9	14.5
	Total	57.1	35.2	7.7	100.0

Notes: Data are weighted. Estimation sample N=1764.

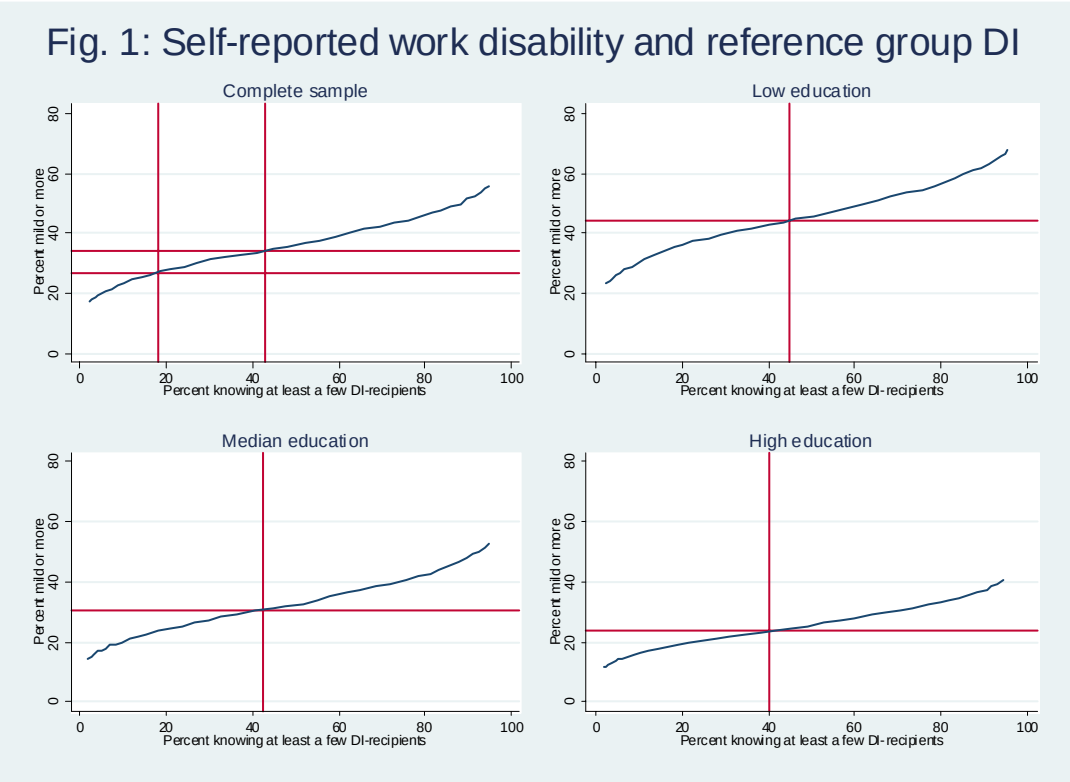
4.1 Simulation of reference group effects

One way to get a feel for the strength of the reference group effects is to artificially vary the number of people on DI in an individual's reference group and show how this affects the prevalence of self reported work limitations. We do this in a very stylized fashion. We vary the intercept in equation (1) and then simulate the reports of DI-benefit receipt in the reference group and the prevalence of self reported work disability. Figure 1 shows the results for both the full sample and broken down by education. In each picture the vertical line represents the percentage of respondents in the sample (or in the relevant education group) who say to know at least a few DI-benefit recipients. The horizontal line represents the percentage reporting to suffer from at least a mild work limitation (either for the full sample or for the relevant

education group). In the picture for the full sample, a second (lower) horizontal line has been drawn. This line is derived from Kapteyn, Smith, Van Soest (2004). They find that if US scales are assigned to Dutch respondents, self-reported work limitations fall by 21%.[6] The second horizontal line is therefore 79% below the first one and may be thought to represent self-reported work-limitations if US scales are applied to Dutch responses.

The picture for the full sample shows that if the percentage of individuals saying they know at least a few DI-benefit recipients in their reference group were to move from its simulated sample mean of 42.9% to about 18.2% (the left most vertical line), this would move the scales used by the respondents enough to get to the US scales. Whether or not this provides a full explanation for the findings by Kapteyn, Smith, and Van Soest (2005) cannot be said without asking the same reference group DI question to an American sample. The level of the graphs for the three education groups varies quite a bit, with the graph for the lowest education group being the highest. That is, at the same level of perceived reference goup DI benefit receipt lower educated respondents are more likely to report at least a mild work limitation than respondents with middle or higher education.

[6] According to their benchmark model; the percentage varies slightly depending on which model specification is chosen.



5 Concluding Remarks

We estimated a stylized model of self reported disability with emphasis on how the reporting is affected by the prevalence of disability in one's reference group. We find an effect in the hypothesized direction, which is rather precisely estimated. The findings are suggestive of how policy affects social norms. In this context, if a policy makes receipt of DI benefits more attractive or easier (e.g., by loosening eligibility requirements) thus increasing the number of DI recipients, this changes social norms. Individuals are now more likely to label a given health condition as work limiting.Thus the prevalence of self-reported work disability will go up.

It is worthwhile to discuss alternative reasons why self-reported disability and reported DI benefit receipt in one's reference group would be correlated and the extent to which our model captures these. First of all let's consider the possibility that individuals with a work disability are more likely to associate with others who suffer a similar fate. Notice that equations (1) and (3) have a considerable number of observable covariates in common, which will generate correlation between self-reported

disability and reported DI benefit receipt in one's reference group. Furthermore, we have allowed for correlation between ω^R and ϵ, although that correlation turned out to be negligible. Next consider the possibility of justification bias, i.e. respondents who are likely to report a work limiting health condition are also more likely to say they know many others on DI. In principle this would be captured by the parameter μ, which relates an individual effect in the response scales for work disability (τ_j) to the response scales for reported DI benefit receipt in the reference group (ϕ_j). Somewhat surprisingly μ is estimated at -.82, which would imply the opposite of justification bias.

In an earlier paper (Kapteyn, Smith, Van Soest, 2004) we have compared self-reported disability in the U.S. and The Netherlands; we find that about 60% of the observed difference in self reports of work disability in the age bracket 51-65 can be explained by response scale differences. In that paper no statement is made about where the difference in scale is coming from. Figure 1 suggests that for our reference group hypothesis to explain the complete difference, Americans should be about half as likely to report that they know at least a few people in their reference group receiving DI benefits. It is suggestive that the actual prevalence of DI benefit receipt in the U.S. is a little less than half of what it is in The Netherlands. Of course, this does not translate directly into a halving of respondents knowing at least a few people on DI. Only new data to be collected for the U.S can tell us whether this is true.

References

[1] James P. Smith Arie Kapteyn and Arthur H.O. van Soest. Vignettes and self-reports of work disability in the us and the netherlands. *American Economic Review*, forthcoming, 2006.

[2] John Bound. Self-reported versus objective measures of health in retirement models. *Journal of Human Resources*, 26(1):106–138, 1991.

[3] John Bound and Richard Burkhauser. Economic analysis of transfer programs targeted on people with disabilities. In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 3C, pages 3417–3528. Elsevier, Amsterdam, 1999.

[4] Leo Aarts Richard Burkhauser Philip de Jong. *Curing the Dutch Disease: An International Perspective on Disability Policy Reform*. Advanced Textbooks in Economics. Aldershot Avebury, 1996.

[5] Joshua Salomon Gary King, Christopher Murray and Ajay Tandon. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1):567–583, 2004.

[6] Robert Havemen and Barbara Wolfe. The economics of disability and disability policy. *Handbook of Health Economics Vol. 1B J. Newhouse and A. Culyer (eds)*, pages 995–1051, 2000.

[7] James P. Smith James Banks, Arie Kapteyn and Arthur H.O. van Soest. Work disability is a pain in the *****, especially in england, the netherlands, and the united states. *RAND Labor and Population Working Paper*, WR-280, April 2005.

[8] Joshua Salomon Ajay Tandon and Christopher Murray. Comparability of self rated health: Cross sectional multi-country survey using anchoring vignettes. *British Medical Journal*, 328(7434):258–260, 2004.

A Vignette Questions

Vignettes for Affect

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

2. [Jim] enjoys work very much. He feels that he is doing a very good job and is optimistic about the future.

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

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

5. [Roberta] feels depressed most of the time. She weeps frequently at work and feels hopeless about the future. She feels that she has become a burden to her co-workers and that she would be better dead.

Vignettes for Pain

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.

Vignettes for CVD

1. [Trish] is very active and fit. She takes aerobic classes 3 times a week. Her job is not physically demanding, but sometimes a little stressful.

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

3. [Paul]'s family has a history of heart problems. His father died of a heart attack when Paul was still very young. The doctors have told Paul that he is at severe risk of having a serious heart attack himself and that he should avoid strenuous physical activity or stress. His work is sedentary, but he frequently has to meet strict deadlines, which adds considerable pressure to his job. He sometimes feels severe pain in chest and arms, and suffers from dizziness, fainting, sweating, nausea or shortness of breath

4. [Tom] has been diagnosed with high blood pressure. His blood pressure goes up quickly if he feels under stress. Tom does not exercise much and is overweight. His job is not physically demanding, but sometimes it can be hectic. He does not get along with his boss very well.

5. [Dan] has undergone triple bypass heart surgery. He is a heavy smoker and still experiences severe chest pain sometimes. His job does not involve heavy physical demands, but sometimes at work he experiences dizzy spells and chest pain.

B Reference Group Questions

The questions are preceded by the following introduction:

The following questions concern your circle of acquaintances, that is, the people with whom you associate frequently, such as friends, neighbors, acquaintances, or maybe people at work.

- If you think of your circle of acquaintances, into which age category do MOST of these people go? Please select the answer that is closest to reality.

age (in years) is mostly:
1 under 16
2 16 - 20
3 21 - 25
4 26 - 30
5 31 - 35
6 36 - 40
7 41 - 45
8 46 - 50
9 51 - 55
10 56 - 60
11 61 - 65
12 66 - 70
13 71 or over

- Which level of education do most of your acquaintances have?

1 primary education
2 junior vocational training
3 lower secondary education
4 secondary education/pre-university education
5 senior vocational training
6 vocational colleges/first year university education
7 university education

- If you think of the men among your acquaintances, how many of them are on DI?

1 Nobody
2 Very few
3 A few
4 Many

- If you think of the women among your acquaintances, how many of them are on DI?

1 Nobody
2 Very few
3 A few
4 Many

C Estimates for threshold and vignette equations

Table C.1: Estimation Results threshold equations

	Threshold shifts					
	γ_1			γ_2		
	Coef.	s.e.	t-value	Coef.	s.e.	t-value
age	0.256	0.091	2.820	-0.055	0.025	-2.169
age squared	-0.022	0.008	-2.633	0.005	0.003	2.140
medium education	0.082	0.037	2.209	-0.049	0.015	-3.215
higher education	0.095	0.037	2.561	-0.033	0.015	-2.251
female	-0.139	0.051	-2.728	0.043	0.013	3.414
stroke	-0.174	0.166	-1.047	-0.051	0.054	-0.938
cancer	-0.061	0.085	-0.719	0.021	0.030	0.692
lung	0.054	0.074	0.733	0.035	0.028	1.257
heart problems	0.019	0.055	0.351	-0.063	0.028	-2.298
highblood	0.019	0.035	0.547	0.010	0.015	0.632
diabetes	0.013	0.071	0.188	0.036	0.032	1.148
emotional problems	-0.039	0.057	-0.671	-0.023	0.021	-1.095
arthritis	0.068	0.058	1.160	-0.019	0.022	-0.885
vision	-0.032	0.077	-0.418	0.020	0.036	0.558
often pain	0.139	0.047	2.937	-0.020	0.015	-1.304
intercept	-0.896	0.253	-3.540	0.130	0.077	1.689
μ	-0.823	0.075	-10.996			
σ_ξ	0.542	0.054	10.110			

Table C.2: Estimates of Vignette Dummies

Vignette	Coef.	s.e.	t-value
1	0		
2	-1.29	0.06	-19.83
3	0.71	0.04	17.88
4	0.54	0.03	16.97
5	1.24	0.06	19.29
6	0.30	0.02	12.39
7	0.76	0.04	19.64
8	1.59	0.08	20.60
9	1.55	0.08	20.41
10	1.35	0.07	20.08
11	-0.95	0.05	-19.71
12	0.68	0.04	18.91
13	1.15	0.06	19.57
14	0.49	0.03	16.43
15	0.79	0.04	18.40
σ_v	0.52	0.02	21.85
δ	0.01	0.02	0.62