Netspar Network for Studies on Pensions, Aging and Retirement

Arthur van Soest

Tatiana Andreyeva

Arie Kapteyn

James P. Smith

Self Reported Disability and Reference Groups

Discussion Paper 2007 - 002

May 29, 2007

Self Reported Disability and Reference Groups*

Arthur van Soest, RAND and Tilburg University†
Tatiana Andreyeva, Yale University and RAND
Arie Kapteyn, RAND James P. Smith, RAND

May 29, 2007

Abstract

Individuals are influenced by the types of people with whom they associate and who form their social networks. These social interactions may affect individual and social norms. We develop a direct test of this using Dutch survey data on how respondents evaluate work disability of hypothetical people with some work related health problem (vignettes). We analyze how the thresholds respondents use to decide what constitutes a (mild or more serious) work disability depend on the number of people receiving disability insurance benefits (DI) in their reference group. To account for endogeneity of DI receipt in a respondent's reference group, we jointly model the respondent's own self-reported work disability, the evaluation thresholds, and DI receipt in the reference group. We find that reference group effects are significant, and contribute substantially to an explanation of why self-reported work disability in the Netherlands is much higher than in, e.g., the US. This implies an important role for social interactions and norms on the perception of work limitations.

*This research was funded by the National Institute of Aging.
†Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands, email: avas@uvt.nl

1 Introduction

In contrast to other social scientists, economists have long adhered to an individualistic notion of behavior, despite early contributions by, for example, Duesenberry (1949) and Veblen (1899). An important modern contribution to the modeling of social interactions is the seminal work of Becker (1974). Although of wider relevance, Becker's work emphasized the interactions among family members, caused by interdependent utilities as well as a common budget constraint. In more recent years, economists have increasingly recognized that individual actions are fundamentally influenced by the attributes and behaviors of those other individuals who form their social networks; see Topa (2001).

The span of behaviors that have been examined in this new research on social interactions has been expanding rapidly and even a very partial list now includes criminal activity (Glaeser, Sacerdote & Scheinkman 1996), (Glaeser, Sacerdote & Scheinkman 2000) neighborhood effects on youth behavior (Case & Katz 1991), models of herd or copycat like behaviors (Banerjee 1992), 'peer effects' in education (Hanuschek, Kian, Markman & Rifkin 2000), (Ginther, Haveman & Wolfe 2000), agglomeration economies (Audretsch & Feldman 1996), information exchanges in local labor markets (Topa 2001), labor supply (Woittiez & Kapteyn 1998), consumption (Kapteyn, van de Geer, van de Stadt & Wansbeek 1997) (Alessie & Kapteyn 1991), retirement plan choices (Duflo & Saez 2003) and social learning through neighbors (Bala & Goyal 1998). As these examples illustrate, the type of social interactions studied has moved well beyond the immediate family to much larger circles of friends, neighbors, and like minded consumers and workers. Various reasons are given for why these types of social interactions matter, including information sharing, demonstration effects, and the formation of tastes and preferences.

Social interactions may also affect what individuals believe to constitute acceptable or normal behavior based on the standards of the sub-communities in which they live and work. In this paper, we develop a direct test of this using data from a household survey representative of the Dutch population on how respondents evaluate work disability of hypothetical people with some work related health problem (vignettes). Combining this with self-reports on the number of people receiving disability insurance benefits (DI) among one's friends and acquaintances, we estimate a model describing the influence of DI prevalence in one's reference group on the subjective scale used to report own and others' work disability.

Both the prevalence of DI benefit receipt and self reported work disability vary substantially across countries; see Haveman & Wolfe (2000) and Bound & Burkhauser (1999). In particular, both are much higher in The Netherlands than in the United

States. Bound & 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. Kapteyn, Smith & van Soest (2007) 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% in The Netherlands against 22.7% in the U.S.). While the higher level of Dutch participation in DI programs is not surprising given higher DI benefits and easier eligibility compared to the US,[1] greater Dutch prevalence of self-reported work disability is puzzling as the Dutch population appears to be healthier than the Americans.[2]

Kapteyn et al. (2007) investigated to what extent differences in self reported work disability can be ascribed to differences in reporting styles and thresholds across countries. Exploiting the vignette methodology originally developed by King, Murray, Salomon & Tandon (2004), Dutch and US respondents were given the same descriptions of work disability problems for hypothetical persons("vignettes"). Dutch respondents appeared to be much more likely to describe the same work disability problem as constituting a work disability than American respondents. Kapteyn et al. (2007) found that more than half of the observed difference in self-reported work disability between the two countries can be explained by differences in response scales.

This result implies that US and Dutch respondents have different norms for evaluating work disability. Our paper analyzes to what extent this is due to peer group effects: do respondents with many DI recipients in their peer group have social norms that make them more likely to evaluate given health problems as constituting a work disability?

We formalize this notion by introducing the concept of prevalence of DI benefit receipt in one's *reference group*, defined as one's circle of friends and acquaintances. In a Dutch survey that we designed and implemented, we asked 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 a "peer group effect," i.e., by the number of people in the reference group receiving disability benefits. To identify the determinants of response scales,

[1] See for instance Aarts, Burkhauser & de Jong (1996). In 2004, DI recipients in The Netherlands made up 13% of the labor force (Source: Statistics Netherlands http://statline.cbs.nl/StatWeb.), while in the US DI-recipients constituted 4.8% of the civilian labor force (Source: US Bureau of Labor Statistics ftp://ftp.bls.gov/pub/news.release/History/empsit.01072005.news)

[2] This is suggested by the analysis of a broad set of health conditions by Banks, Kapteyn, Smith & van Soest (2007).

we exploit anchoring vignettes as in Kapteyn et al. (2007). Using this additional information helps to solve the identification problem that is present in many models with peer group effects, known as the reflection problem (Manski (1993)).

The remainder of the paper is organized as follows. 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. One equation explains the answers to the question about DI benefit receipt in the respondents' reference group. A 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 summarizes our main results. 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. The final section presents our conclusions.

2 The Data

In this research, we use information obtained from the Dutch CentERpanel. This is an Internet panel of about 2,250 households who have agreed to respond to a survey every weekend. 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 not restricted to households with Internet access, but representative of the Dutch adult population except the institutionalized. 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 characteristics of the respondent and household, including 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 October 2003, we fielded a second wave of vignettes with slightly different wording of the questions, and also included questions about reference groups. For our analysis we will use the vignette and reference group data from this October wave. Appendix A lists the vignette questions. All vignettes are presented with either a female or a male name.[3]

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. The differences in distributions of answers correspond quite well with the variation in severity of the conditions described in the vignettes. For example, in all three domains of affect, pain, and CVD, the condition described in the third vignette seems much more severe than that described in the first, and respondents ranked them accordingly. Moreover, there was also a great deal of consistency among respondents in how they ordered vignettes in terms of their severity, showing that respondents understood these experiments and took their responses seriously; see Banks et al. (2007) for details.

Table 2 presents the distribution of the answers to the question on own work limitations by age group. These represent answers to the question:

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

The question allows respondents to reply on the 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.

Table 2 implies that about 37% of the Dutch population reports to have at least a mild work limitation and about 14% have a work limiting health problem or impairment that they gauge as moderately limiting or worse. Not surprisingly, work related health deteriorates with age (although cohort effects may also play some role in this pattern).

[3]Female or male names are assigned randomly. In Appendix A we only show one of the two names per vignette.

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 A for the wordings of the vignette questions.

The most interesting groups are probably people in the age groups 45-54 and 55-64. For them, the prevalence of work limiting health problems is large, and this will often be an important reason not to participate in the labor market. For the 65-plus, work limiting health problems are even more prevalent, but these people are almost always retired anyhow.

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 and 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 in the respondent's reference group. In the analysis we will combine the age and education categories into a smaller number of broader brackets. Table 3 presents descriptive statistics for our independent variables, 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 acquaintances have a medium education level (while 39 percent of the respondents has that level).

The other reference group questions refer to the number of acquaintances receiving disability benefits, separately for men and women. These are the crucial variables for our analysis as they measure DI benefit receipt 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. We discuss the sensitivity of our results to this definition of the reference group variables below.

The distribution of reported DI receipt in the reference group 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", because the frequencies for "Few" and particularly "Many" are small. Young people typically know no one 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 a government provided pension and often one or more additional occupational pensions. 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 (e.g. to raise children) and never qualify for disability benefits after that.

Plausibly, these reference group variables are endogenous to the respondent's own work disability – respondents who have a work disability will often not work and will not only receive disability benefits, but will also more easily get acquainted with other people on disability benefits. Hence we will treat the number of acquaintances on disability benefits as a dependent variable, modelled jointly with work limitations. Table 5 shows cross tabs of self-reported work limitations and self-reported prevalence of DI-receipt in one's reference group. For simplicity of presentation, 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.

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. Second, 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). Third, there may be other (observed or unobserved) factors that both increase the likelihood that respondents have a work limitation and that they know others with work limitations. One such factor is age. Fourth, response scales used in answering the reference group questions might be correlated with response scales in self-reported work disability. Respondents may, for instance, exaggerate the number of friends or acquaintances on DI to "justify" their own report of a work limitation (Bound 1991). These explanations are not mutually exclusive and undoubtedly there are more. We are particularly interested in the role played

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 (Groningen, Friesland & Drenthe)	14.3
Eastern provinces (Overijssel, Flevoland & Gelderland)	21.6
Western provinces (Utrecht, Noord-Holland & Zuid-Holland)	38.7
Southern provinces (Zeeland, Noord-Brabant & Limburg)	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
Education level in the reference group: Low	24.9
Education level in the reference group: Medium	47.9
Education level in the reference group: High	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.

Table 4. Distribution of DI Receipt in 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.

by the first explanation, reflecting a social interaction effect. In the next section we present a model that aims at isolating the importance of the first explanation; in the discussion of the results we will also return to the competing explanations.

3 A Model with Reference Groups

Our econometric model explains the reported number of people on disability in the reference group R (cf. Table 4), self-reported work disability Y (cf. Table 2), and disability of vignette persons $Y^1, \ldots, Y^{15}$ (cf. Table 1).

Self-reports of own work disability

Individuals evaluate the extent of their work disability with a self-evaluation of whether their health problems and working conditions are sufficiently problematic to place them above their own subjective threshold of being somewhat limited or more than somewhat limited. The result of that evaluation depends on the extent of their true health problems as well as their subjective thresholds of what constitutes a disability, both of which vary across individuals.

Table 5. Self-Reported Work Disability and DI Receipt in Reference Group

		DI Receipt in the Reference Group, %			
		None	Very few	A few / many	Total
Self-reported work disability, %	Not limited	63.4	32.2	4.4	100.0
		68.1	58.1	38.0	62.5
	Mildly limited	51.9	38.4	9.7	100.0
		20.8	25.9	30.9	23.3
	Moderately, severely and extremely imited	45.3	38.7	16.0	100.0
		11.1	16.0	31.1	14.2
	Total	58.1	34.6	7.3	100.0
		100.0	100.0	100.0	100.0

Notes: Data are weighted. Estimation sample N=1764.
Top rows: row percentages; Bottom rows: column percentages

More formally, self-reported work disability Y of respondent i is modeled on a 3-point scale of not at all limited, somewhat limited, and more than somewhat limited (combining the three most serious categories "moderate," "severe," and "extreme," to one) as follows:

$$Y_i^* = X_i\beta + \epsilon_i \tag{1}$$

$$Y_i = j \text{ if } \tau_i^{j-1} < Y_i^* \leq \tau_i^j,\ j = 1, 2, 3. \tag{2}$$

For notational convenience, we define $\tau_i^0 = -\infty$ and $\tau_i^3 = \infty$. The remaining thresholds τ_i^1 and τ_i^2 will be modeled as functions of observable and unobservable respondent characteristics as described below. The error term ϵ_i is assumed to be standard normally distributed. (Complete assumptions on error terms are given below.)

Since thresholds depend on respondent characteristics, self-reported work disability alone is not enough to distinguish between variation in Y_i^* and variation in the thresholds, i.e., genuine variation in work related health and variation in what constitutes a disability in respondents' perceptions. Vignettes are used to identify this distinction.

Vignette evaluations

The vignettes provide all respondents with the descriptions of the same set of work disability problems. As a consequence, variation in how respondents evaluate the given health problems informs us about variation in the subjective thresholds used by the respondents. More formally, the evaluations Y_i^l of vignettes l, $l = 1, \ldots, 15$, are given by

$$Y_i^{l*} = \theta^l + \delta F_i^l + \epsilon_i^l \tag{3}$$

$$Y_i^l = j \text{ if } \tau_i^{j-1} < Y_i^{l*} \leq \tau_i^j, \ j = 1, 2, 3. \tag{4}$$

Here F_i^l is a dummy variable indicating whether the person described in the vignette is female ($F_i^l = 1$) or not ($F_i^l = 0$). This specification follows earlier work by Kapteyn et al. (2007), who find that respondents (both males and females) tend to be "harsher" on female than on male vignette persons, i.e., $\delta < 0$. We assume that all ϵ_i^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_i^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 not be captured by ϵ_i^l but by an unobserved heterogeneity term in the response scales, see below.

Response scale thresholds

The crucial assumption guaranteeing that vignettes help to identify response scale differences, is that individuals use the same scales in evaluating themselves as they do with the vignette persons (response consistency, cf. King et al. (2004). The thresholds used in the vignette evaluation can vary across all types of individual attributes. In this study, we expand the set of attributes and include the number of persons among friends and acquaintances who are on disability benefits R_i^*. The thresholds τ_i^1 and τ_i^2 are modeled as follows:

$$\tau_i^1 = V_i\gamma_1 + \gamma_1^R R_i^* + \xi_i \tag{5}$$

$$\tau_i^2 = \tau_i^1 + e^{V_i\gamma_2 + \gamma_2^R R_i^*} \tag{6}$$

We have included a vector V_i of respondent characteristics (independent of all error terms and not including the reference group variables) 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 key parameters

of interest are γ_1^R and γ_2^R, the estimated impact of the number of people on DI in one's reference group on the threshold that is used to evaluate work disability. 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_i reflects unobserved heterogeneity in thresholds. For computational convenience, we do not allow for unobserved heterogeneity in the distance between the two thresholds. ξ_i is assumed to follow a normal distribution with variance σ_ξ^2, independent of V_i and other error terms in the model.

DI receipt in the reference group

As explained above, we consider DI receipt in the respondent's reference group of the respondent's own sex and 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, $j = 1$ ("none"), $j = 2$ ("very few") and $j = 3$ ("a few" or "many"). This will be modeled with an ordered probit equation:

$$R_i^* = X_i^R \beta^R + \omega_i^R, \quad \omega_i^R \sim N(0, \sigma_\omega^2) \tag{7}$$

$$R_i = j \text{ if } \phi_i^{j-1} < R_i^* \leq \phi_i^j, \ j = 1, 2, 3. \tag{8}$$

For notational convenience, we define $\phi_i^0 = -\infty$ and $\phi_i^3 = \infty$. Below we will further specify the thresholds ϕ_i^1 and ϕ_i^2. The vector X_i^R of respondent characteristics driving reference group disability is assumed to be independent of all the errors in the model. Equation (7) has a "reduced form" nature in the sense that we do not explicitly model how work disability and labor force status affect disability in the reference group. The exogenous determinants of labor force status and disability are included among the regressors X_i^R to account for this.

Since it is likely that there are common unobserved factors affecting both the number of people one knows on disability benefits and one's own evaluation of work disability, we allow for correlation between ϵ_i and ω_i^R. This correlation also allows 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 composition of the reference group.

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

We allow for a common unobserved heterogeneity component driving the thresholds $\tau_i^j, j = 1, 2$ and the thresholds in the reference group equation $\phi_i^k, k = 1, 2$ by specifying: $\phi_i^1 = \phi_0^1 + \mu\xi_i$ and $\phi_i^2 = \phi_0^2 + \mu\xi_i$. We normalize $\phi_0^1 = 0$. The parameter μ could be positive (respondents exaggerating their work disability also exaggerate their number of acquaintances on DI) or negative (respondents who think of work disability as something exceptional will tend to interpret a given number of acquaintances on DI as large).[5] ϕ_{02} and μ are additional parameters to be estimated. Define $u_i^R = \omega_i^R - \mu\xi_i$ and, for notational convenience, let $\phi_0^0 = -\infty$ and $\phi_0^3 = \infty$. For normalization we set $\text{Var}(u_i^R) = 1$. We can then rewrite (8) as

$$R_i = j \text{ if } \phi_0^{j-1} < X_i^R\beta^R + u_i^R \leq \phi_0^j, \ j = 1, 2, 3. \tag{9}$$

Error terms and identification

The error terms in the model, including unobserved heterogeneity components, are: $\epsilon_i, \epsilon_i^l, l = 1, ..., 15, \omega_i^R$, and ξ_i. We assume they are all normally distributed and independent of the regressors X_i, X_i^R, V_i and F_i^l. The only correlation we allow for is between ϵ_i and ω_i^R. We assume (ϵ_i, ω_i^R) is bivariate normal with correlation coefficient ρ. The assumption that ξ_i is independent of ϵ_i 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_i and V_i constant. 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

Like in all models with reference group effects, identifying the causal effect of the reference group variable requires model assumptions, due to endogeneity issues and confounding effects (cf. Manski 1993). To see what are the main identifying assumptions in the model described above is difficult due to non-linearities. Because of this, we present a stripped down linear version of the model in Appendix C, in which the identification (other than through non-linearities) issue is essentially the same.

In this model, the various confounding effects discussed by Manski are addressed. We argue that identifying assumptions are more plausible than in Manski's standard case, mainly because we have direct information on the reference group variables. Moreover, the dependent variable (work disability) is not the same as the variable of which the reference group mean is taken (receiving disability benefits). Finally,

[5] It seems natural to add another error term to the ϕ_i^j which is independent of everything else, but this will be subsumed in ω_i^R.

a distinction can be made between scale shifts and genuine shifts due to the use of vignettes.

4 Results

We estimate the models using simulated maximum likelihood. Appendix D contains the likelihood according to the model formulated above. The integrals in the likelihood contributions ((24) in Appendix D) are replaced by smooth simulation-based approximations, by drawing 200 times from the joint distribution of ξ and u^R and using Halton draws.[6] Experiments with a substantially larger number of draws did not lead to appreciable differences in the results, implying that the number of draws is large enough to provide an accurate approximation of the integral.

4.1 Estimation results

Table 6 presents the estimation results for self reported work disability and for DI receipt in the reference group (equations (1) and (7)). The estimates for the threshold equations (5) and (6) are given in Table 7. Estimates for the vignette equations (3) are not of primary interest and are therefore presented Table E.1 in Appendix E, where also a brief discussion is given.

Work disability self-reports

The equation for self-reported disability in Table 6 shows that self-reported disability goes up with age until age 69; 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.

DI receipt in the reference group

The reference group DI receipt equation shows that the reported DI prevalence in the reference group increases with age (the top of the parabola is estimated at 82 years); it shows virtually no relation with 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. Also in line with the raw data (Table 4) is that females are less

[6] We have used the program mdraws written by Lorenzo Cappellari and Stephen P. Jenkins. See Cappellari & Jenkins (2006)

Table 6. Estimation Results for Own Work Disability and DI Receipt in Reference Group

	Own Work Disability		DI Receipt Ref. group	
	Coef.	**s.e.**	**Coef.**	**s.e.**
age	0.290	0.146	0.326	0.170
age squared	-0.021	0.014	-0.020	0.016
medium education	-0.063	0.084	0.076	0.076
higher education	-0.273	0.085	-0.049	0.083
female	0.026	0.070	-0.339	0.063
stroke	1.304	0.340	0.070	0.130
cancer	0.337	0.146	0.136	0.153
lung	0.617	0.133	0.185	0.172
heart problems	0.872	0.131	0.022	0.182
highblood	0.037	0.084	-0.166	0.201
diabetes	0.403	0.178	0.006	0.062
emotional problems	0.625	0.097	-0.029	0.075
arthritis	0.423	0.118	-0.016	0.081
vision	0.098	0.178	0.019	0.071
often pain	1.231	0.081	-0.158	0.064
intercept	-2.005	0.376	-0.064	0.257
reference group age 25-35			-0.177	0.160
reference group age 36-45			0.295	0.134
reference group age 46-55			0.003	0.119
reference group age 56-65			0.065	0.075
reference group age >65			0.155	0.153
Medium education in R.G.			0.268	0.099
High education in R.G.			0.197	0.109
northern provinces			0.041	0.166
eastern provinces			0.252	0.078
western provinces			-1.250	0.388
ρ	0.055	0.039		
φ_{02}	1.336	0.051		

likely to report to have 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 variables affecting the number of people on DI in the reference group are of interest in part because, as we shall see below, the number of people in the reference group significantly affects the thresholds used in evaluating work disability. For example, women know fewer people on DI and because of that will less easily say that a given health problem constitutes a work disability. Similarly, having pain increases the number of people on DI in one's reference group, and this makes people with pain 'softer' in evaluating disability.

Thresholds

The results for the threshold equations are presented in Table 7. The top panel presents estimates for the coefficients on individual characteristics in equations (5) and (6), while the bottom part shows estimates of the coefficients of peer group DI receipt R_i^* interacted with education, age, and gender in both threshold equations. The mean value of R_i^* is positive. Thus the estimates for the first threshold imply that women with the mean disability receipt in their peer group, use lower thresholds than similar men, and thus more easily regard a health problem as work limiting. People with higher education are less likely to evaluate a given health problem as work limiting than low educated respondents. The age function has a top at about 62 years, implying that until age 62, older people are "tougher", i.e. less likely to call a condition work disabling.[7] The only significant health condition is pain - respondents who often suffer from pain less easily evaluate a given health problem as a (mild or worse) work limitation.

For the distance between the first and second threshold (γ_2), results are different; the age function has a minimum at 54 years of age (if $R_i^* = 0$), while higher education leads to a smaller distance between thresholds. Heart problems do the same; they are the only type of health problems with a significant effect.The estimates are difficult to interpret individually, due to the complexity of the model, where the same variables appear in several equations.

The parameters of primary interest are γ_1^R and γ_2^R. Both have been parameterized as a function of education level, age, and gender (see the bottom panel of Table 7). Consider first the estimated main effect and the interactions with education. For males under 35 with lower education γ_1^R is estimated at -.45; for the same individuals with medium education the estimate is -.42 (not significantly different from the -

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

Table 7. Estimation Results Threshold Equations

	Threshold shifts			
	γ_1		γ_2	
	Coef.	**s.e.**	**Coef.**	**s.e.**
age	0.234	0.098	-0.076	0.028
age squared	-0.019	0.009	0.007	0.003
medium education	0.071	0.039	-0.045	0.016
higher education	0.080	0.038	-0.032	0.015
female	-0.110	0.051	0.021	0.014
stroke	-0.100	0.154	-0.051	0.053
cancer	-0.061	0.086	0.023	0.031
lung	-0.056	0.076	0.031	0.028
heart problems	0.056	0.062	-0.062	0.028
highblood	0.003	0.039	0.020	0.016
diabetes	0.000	0.079	0.025	0.032
emotional problems	0.001	0.059	-0.023	0.021
arthritis	0.039	0.059	-0.016	0.022
vision	-0.057	0.082	0.031	0.036
often pain	0.100	0.048	-0.014	0.015
intercept	-0.837	0.276	0.178	0.085
μ	-0.816	0.086		
σ_ξ	0.523	0.059		
Interactions	γ_1^R		γ_2^R	
	Coef.	**s.e.**	**Coef.**	**s.e.**
age35-64	0.037	0.030	-0.059	0.021
age65+	0.012	0.048	-0.009	0.027
medium education	0.029	0.029	-0.066	0.019
higher education	0.118	0.031	-0.052	0.018
female	0.063	0.026	-0.073	0.017
intercept	-0.451	0.133	0.166	0.027

0.45 estimate), while for the higher educated the estimate is -.33. The difference between the highest and lowest education groups is statistically significant. Females are significantly less influenced by DI receipt in the reference group than males, although quantitatively the difference is not big. Interactions of DI receipt in the reference group with age dummies are small and insignificant.

As long as the estimated value of γ_1^R is negative (as it is in all cases), the fraction of people who are on DI benefits in the reference group will unambiguously shift the reporting thresholds for own disability downward. In this sense, γ_1^R is the more critical parameter of the two. The estimates for γ_2^R show that the distance between thresholds increases with the number of friends and acquaintances on disability benefits, particularly for the lowest education level and for age less than 35 or above 65.

In simulations, we find that if the number of people on DI in the reference group increases, this raises both the fraction of those reporting they are somewhat limited and the fraction of those reporting they are moderately limited or worse. As mentioned earlier, we defined reference groups separately for men and women in the sense that for women we took the number of women on DI amongst female acquaintances and for men the number of male DI recipients among male acquaintances. One question is how sensitive our results in Table 7 are to this particular formulation of reference groups. To test this, we re-estimated the model using a common definition of reference groups for both sexes.[8] The estimated effects of number of people on DI in the reference group are even larger using the common reference by gender than with the benchmark definition used for Table 7. A likelihood ratio test however indicates that the model we estimate with separate reference groups by gender is significantly better that the alternative model.

Covariance structure of the errors

Table 6 shows that the parameter ρ, the correlation between the error terms in the equations for own work disability (1) and DI receipt in the reference group (7) is small and insignificant. Unobserved heterogeneity in thresholds is significant - the estimated standard deviation of ξ is 0.52 and seems to be very accurately determined (σ_ξ in Table 7). To judge its size, it can be compared to the amount of idiosyncratic noise in self-reports and vignette evaluations. The former has standard

[8] All respondents were asked both the numbers of men and the numbers of women on DI in the reference group. To form a common definition, we used the maximum of the two. Thus if for an individual respondent there were a lot of individuals of one gender who were more than somewhat limited, that is the value that applies.

deviation 1 (by normalization), the latter has standard deviation 0.51 (see Table E.1). Thus unobserved heterogeneity in thresholds explains about 23% of the unsystematic variation in self-reports and about 53% of the unsystematic variation in vignette evaluations.

The parameter μ is estimated at -0.82. Since $u^R = \omega^R - \mu\xi$ and $\text{Var}(u_i^R) = 1$, we have $\text{Var}(\omega_i^R)$=0.82. The implied correlation between ξ_i and u_i^R is equal to -0.43. The sign of μ 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. Thus someone who is unlikely to refer to a health problem as work limiting, will sooner consider a given number of people on DI in the reference group as "many".

4.2 Model performance

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 Table 8 with Table 5 suggests that the fit of the model is fairly good; judging from the marginal distributions, 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 19.6% in that category.

4.3 Simulation of reference group effects

One way to gauge the strength of the reference group effects is to artificially vary the number of people on DI in an individual's reference group and then to evaluate how this affects the prevalence of self reported work limitations. We do this by varying the intercept in the equation for the number of people in the reference group on DI (7) and then simulate the reports of DI-benefit receipt in the reference group and the prevalence of self reported work disability induced by that new level of reference group DI-receipt.

Figure 1 shows the results for both the full sample and the sub-samples broken down by education. In each picture the horizontal axis is the percentage of respondents who say that they know at least a few DI-benefit recipients, with the vertical lines representing the sample (or subsample) percentages (except the left vertical line in the first figure, see below). The vertical axis represents the percentage who report

Table 8. Model Predictions of Self-Reported Work Disability and DI Receipt in Reference Group

		Disability in the reference group, %			
		None	Very few	A few / many	Total
Self-reported work disability, %	Not limited	62.1	32.3	5.7	100.0
		71.8	60.4	48.7	66.0
	Mildly limited	51.4	38.4	10.3	100.0
		17.6	21.3	26.2	19.6
	Moderately, severely and extremely limited	42.1	44.5	13.4	100.0
		10.6	18.2	25.1	14.4
	Total	57.1	35.2	7.7	100.0
		100.0	100.0	100.0	100.0

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

Top rows: row percentages; Bottom rows: column percentages

that they suffer from at least a mild work limitation; the horizontal line indicates the (sub)sample percentage (except the lower line in the first figure).

The curves in the figures illustrate the sensitivity of reporting a work disability to DI receipt in the reference group. In line with the estimation results in Table 7, both level and slope are largest for the low educated and smallest for those with a high education level. The difference in level implies that 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.But in all cases there is a notable effect - if the respondent knows more people on DI benefits, the chances of reporting a disability increase substantially.

To illustrate the size of the effect, in the picture for the full sample, additional horizontal and vertical lines have been drawn, both below the sample averages. The horizontal line is based on the finding of Kapteyn et al. (2007) that if US scales are assigned to Dutch respondents, self-reported work limitations in the Netherlands would fall by 21%.[9] This second horizontal line can thus be interpreted as self-reported work-limitations in the Netherlands if US scales are applied to Dutch

[9]This is the finding in their benchmark model; the percentage varies slightly depending on which model specification is chosen.

responses. The second vertical line 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 reach the US scales.

5 Concluding Remarks

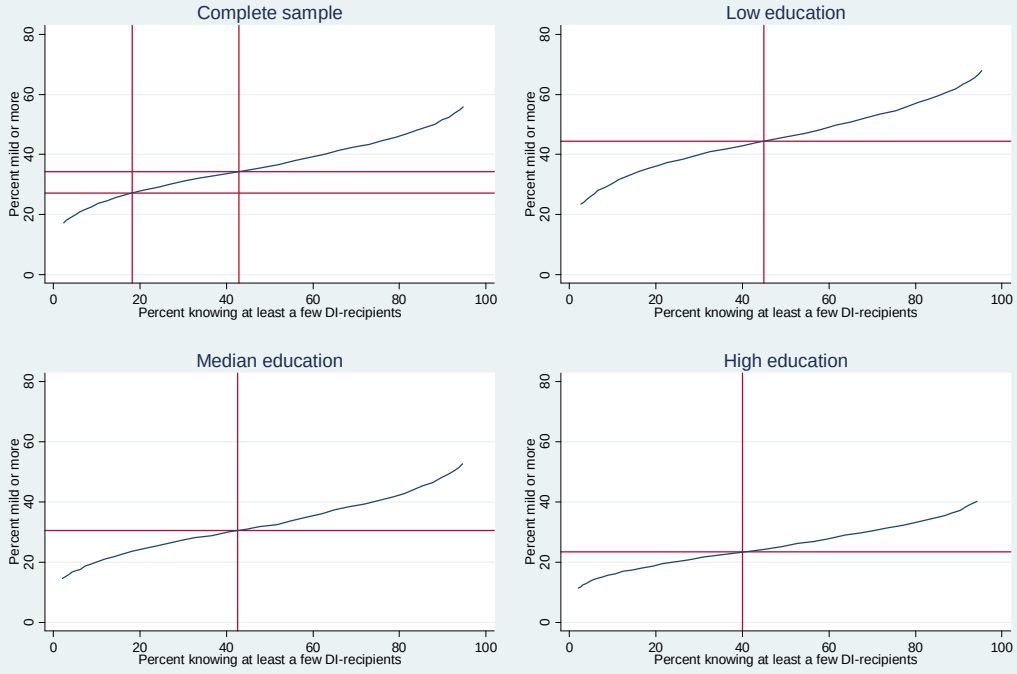
Most people do not live in social isolation. Instead, they interact repeatedly with family, friends, and neighbors. As a consequence of those pervasive interactions, they allow themselves to be transformed in many ways, a transformation of which they may often be unaware. One type of transformation involves the formation of social norms about what normal or acceptable behavior might be. These social norms then fix the thresholds that they may be using in responding to questions about their own behaviors and current situations. If they had different neighbors and friends, their self-descriptions about their lives may well be quite different. While this may be true within a country where there exists a shared history and culture, it is especially likely to be the case when cross-national comparisons are made.

In this paper, we test the importance of these types of social interactions using a specific application- the probability that people self-label themselves as work disabled. We estimated a model of self reported disability with an emphasis on how the reporting of disability is affected by the prevalence of DI receipt in one's reference group. We find an effect in the hypothesized direction- larger reported numbers of people in one's reference group on DI increase the likelihood of seeing oneself as having a work disability.

These findings are suggestive of how policy programs affect social norms. 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 and the prevalence of self-reported work will rise.

There are of course alternative reasons why self-reported disability and reported DI benefit receipt in one's reference group would be correlated. Our model is designed to capture many of these reasons. These include the possibility that individuals with a work disability are more likely to associate with others who suffer a similar fate. First, we allow for a considerable number of observable covariates in common, which by itself will generate correlation between self-reported disability and reported DI benefit receipt in one's reference group. But we also allow for correlation between the errors in the reference group equation and the equation predicting the probability that one is work disabled.

Fig. 1: Self-reported work disability and reference group DI



Even within this reasonably general model, we find a direct effect of the number of people in one's reference group on disability programs on the probability one considers oneself work disabled. The effects that we estimate are sufficiently strong that they are able to explain a good deal of the higher rates of self-reported work disability in the Netherlands compared to the United States. The Dutch population appears to have much more lenient thresholds about what constitutes a work disability (Kapteyn et al. 2007). The results in this paper suggest that this tendency stems from the fact that the Dutch are much more likely to know people on work disability programs, a direct consequence of the far more generous programs in The Netherlands as well as its more lenient rules for program eligibility.

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.

References

Aarts, L., Burkhauser, R. & de Jong, P. (1996), *Curing the Dutch Disease: An International Perspective on Disability Policy Reform*, Advanced Textbooks in Economics, Avebury: Aldershot.

Alessie, R. & Kapteyn, A. (1991), 'Habit formation, interdependent preferences and demographic effects in the almost ideal demand system', *Economic Journal* **101**, 404–419.

Audretsch, D. & Feldman, M. (1996), 'R and d spillovers and the geography of innovation and production', *American Economic Review* **86**, 630–640.

Bala, V. & Goyal, S. (1998), 'Learning from neighbours', *Review of Economic Studies* **65**, 595–622.

Banerjee, A. V. (1992), 'A simple model of herd behavior', *Quarterly Journal of Economics* **107**, 797–817.

Banks, J., Kapteyn, A., Smith, J. P. & van Soest, A. (2007), Work disability is a pain in the *****, especially in england, the netherlands, and the united states, *in* D. Cutler & D. Wise, eds, 'Health in Older Ages: The Causes and Consequences of Declining Disability Among the Elderly', University of Chicago Press, Chicago.

Becker, G. S. (1974), 'A theory of social interactions', *Journal of Political Economy* **83**(6), 1063–1093.

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

Bound, J. & Burkhauser, R. (1999), Economic analysis of transfer programs targeted on people with disabilities, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3C, Elsevier, Amsterdam, pp. 3417–3528.

Cappellari, L. & Jenkins, S. (2006), Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *Discussion Paper* No. 2112, IZA, Bonn.

Case, A. & Katz, L. (1991), The company you keep: The effects of family and neighborhood on disadvantaged youth. *NBER Working Paper 3705*, National Bureau of Economic Research, Cambridge, MA.

Duesenberry, J. (1949), *Income, Saving and the Theory of Consumer Behavior*, Cambridge: Harvard University Press.

Duflo, E. & Saez, E. (2003), 'The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment', *Quarterly Journal of Economics* **118**(3), 815–842.

Ginther, D., Haveman, R. & Wolfe, B. (2000), 'Neighborhood attributes as determinants of children's outcomes: How robust are the relationships', *Journal of Human Resources* **35**(4), 603–642.

Glaeser, E. L., Sacerdote, B. & Scheinkman, J. A. (1996), 'Crime and social interactions', *Quarterly Journal of Economics* **111**(2), 507–548.

Glaeser, E. L., Sacerdote, B. & Scheinkman, J. A. (2000), 'The social multiplier', *Journal of the European Economic Association* **1**(2-3), 345–353.

Hanuschek, E., Kian, J., Markman, J. & Rifkin, S. (2000), Do peers affect student achievement. *Unpublished Manuscript* University of Rochester.

Haveman, R. & Wolfe, B. (2000), 'The economics of disability and disability policy', *Handbook of Health Economics Vol. 1B J. Newhouse and A. Culyer (eds)* pp. 995–1051.

Kapteyn, A., Smith, J. P. & van Soest, A. H. (2007), 'Vignettes and self-reports of work disability in the u.s. and the netherlands', *American Economic Review* **97**(1), 461–473.

Kapteyn, A., van de Geer, S., van de Stadt, H. & Wansbeek, T. (1997), 'Interdependent preferences: An econometric analysis', *Journal of Applied Econometrics* **12**(6), 665–686.

King, G., Murray, C., Salomon, J. & Tandon, A. (2004), 'Enhancing the validity and cross-cultural comparability of measurement in survey research', *American Political Science Review* **98**(1), 567–583.

Manski, C. F. (1993), 'Identification of endogenous social effects', *The Review of Economic Studies* **60**(3), 531–542.

Topa, G. (2001), 'Social interactions, local spillovers and unemployment', *The Review of Economic Studies* **68**(2), 261–295.

Veblen, T. (1899), *The Theory of the Leisure Class*, New York: Dover.

Woittiez, I. & Kapteyn, A. (1998), 'Social interactions and habit formation in a model of female labour supply', *Journal of Public Economics* **70**, 185–205.

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 Identification and the reflection problem

To explain what identifies our model, this appendix presents a simplified linear version with one threshold and one vignette question (for convenience, the same for all respondents, so that the vignette gender dummy is not included). Furthermore we take reference group DI receipt to be observable (rather than being a latent variable that is only indirectly observable). These simplifications do not fundamentally affect identification, but make intuition and comparability with Manski's reflection problem (see Manski 1993) easier. With W_i denoting all individual characteristics (exclusion restrictions will be discussed below), the simplified linear model is given by:

Work disability self-report:

$$Y_i = W_i\beta + \epsilon_i \tag{10}$$

Threshold (shift in all disability reports):

$$\tau_i = W_i\gamma + \gamma^R R_i + \xi_i \tag{11}$$

Reference group DI:

$$R_i = W_i\beta^R + \omega_i^R \tag{12}$$

Reported work disability of vignette person:

$$Y_i^V = \theta - \tau_i + \epsilon_i^V \tag{13}$$

First note that our social group variable is DI receipt in the reference group, but the dependent variable in (1) is not own DI receipt but self-reported own work disability. This is different from Manski's case where the social group variable is essentially a group mean of the individual left hand side variable.

Second, we observe that (12) is already a reduced form equation, i.e. we do not model the causal effect of, for example, own work disability on DI receipt in the reference group (e.g. through DI receipt and socializing with others on DI receipt). However, since it is likely that such a mechanism exists, we definitely want to allow for correlation between the error terms ϵ_i and ω_i^R.

Next, note that (11) excludes one of the effects discussed by Manski: it assumes that reference group means of W_i (such as average age, health, or education of the reference group) do not affect τ_i or Y_i directly, but only through their possible effect on R_i. This seems plausible in our case; we see no reason why there would be a direct effect of such reference group averages on either genuine work disability of the individual, or the individual's reporting scale.

The second confounding effect Manski considers is correlation between unobserved heterogeneity in (10) or (11) and unobserved heterogeneity in (12). As already mentioned, we allow for correlation between ω_i^R and ϵ_i. The need to allow for correlation between ω_i^R and ξ_i seems less obvious, since the dependent variables in (11) and (12) refer to different things. Still, one could argue that those who are "soft" in the sense that they think a small health problem is already a serious limitation, tend to socialize with people drawing (disability) benefits. Therefore, we do allow for a correlation between ω_i^R and ξ_i. (In the model of the main text, this correlation comes in through the term $\mu\xi_i$ in the ϕ_i^j.)

The third effect Manski considers is the causal effect of reference group DI on the variable of interest τ_i. This is the parameter of interest γ^R in (11).

There seems to be no good reason to expect a correlation between ϵ_i^V and any of the other error terms, or between ϵ_i and ξ_i. However, for identification purposes this does not really matter, as we will see below.

Consider the model with the assumptions discussed above. Since the threshold is not observed, we eliminate it from the model. This gives:

Work disability self-report:

$$Y_i = W_i(\beta - \gamma) - \gamma^R R_i + \epsilon_i - \xi_i \tag{14}$$

Vignette disability:

$$Y_i^V = \theta - W_i\gamma - \gamma^R R_i + \epsilon_i^V - \xi_i \tag{15}$$

Using (12, (14) and (15), we obtain the following reduced form:

Work disability self-report:

$$Y_i = W_i(\beta - \gamma - \gamma^R\beta^R) + \epsilon_i - \xi_i - \gamma^R\omega_i^R \tag{16}$$

Reference group DI receipt:

$$R_i = W_i\beta^R + \omega_i^R \tag{17}$$

Vignette disability:

$$Y_i^V = \theta - W_i(\gamma + \gamma^R\beta^R) + \epsilon_i^V - \xi_i - \gamma^R\omega_i^R \tag{18}$$

This leads to the following observations:

- β^R is identified from (17). (This is no surprise since β^R has reduced form coefficients.)

- $\gamma + \gamma^R \beta^R$ is identified from (18) and hence β can be obtained from (16).
- In order to separately identify γ and γ^R there are two options:
  - If some of the W_i appear in (17) but not in (11), then we can separately identify γ and γ^R from (18). This is the case of exclusion restrictions.
  - Alternatively, if we were willing to assume there is no correlation between ω_i^R and ξ_i, then there is no endogeneity problem in (15) and we can estimate γ and γ^R directly using (15).

  As explained above, we have chosen the first option.[10]
- One can see from (16)-(18) that allowing for correlation between ϵ_i^V and any of the other error terms, or between ϵ_i and ξ_i does not affect identification of the structural parameters.

Thus identification of the reference group effect requires exclusion restrictions - variables that affect DI receipt in the reference group, but do not have a direct effect on the evaluation threshold. We use two types of variables: directly elicited reference group variables and regional indicators. These variables will still affect response scales (as represented by T_i), but only through the reference group variable R_i.

We believe these exclusion restrictions to be reasonable. For instance, The Netherlands is a small homogeneous country, where there is no reason to believe there are strong unobserved differences (after controlling for the variables that appear in the threshold equations) that are purely determined by the region one lives in. There are however regional differences in the prevalence of DI receipt, for instance due to differences in industrial structure or generosity of the local organizations that decide about benefit applications. The assumption that the typical age and education in the reference group has no effect on reporting thresholds was already discussed above. These variables can be expected to affect work disability in the reference group.

D Likelihood Contributions

Compared to the models in (King et al. 2004) and (Kapteyn et al. 2007), there are two complications: the thresholds now depend on an unobserved variable R^* and

[10] We have estimated the model also under the alternative assumption. This gives a significant but smaller value of the reference group effect.

upon an unobserved heterogeneity term ξ. Replacing R^* using (7) and exploiting (5) and (6) gives:

$$\tau_1 = V\gamma_1 + \gamma_1^R X^R \beta^R + \xi + \gamma_1^R(u^R + \mu\xi), \tag{19}$$

$$\tau_2 = \tau_1 + e^{V\gamma_2 + \gamma_2^R X^R \beta^R + \gamma_2^R (u^R} + \mu\xi) \tag{20}$$

(1) and (2) imply

$$Y = j \text{ if } \tau_{j-1} - X\beta < \epsilon < \tau_j - X\beta \tag{21}$$

Similarly, for the vignette evaluations we get:

$$Y^l = j \text{ if } \tau_{j-1} - \theta^l - \delta F^l < \epsilon^l < \tau_j - \theta^l - \delta F^l \tag{22}$$

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

$$R = j \text{ if } \phi_{0,j-1} - X^R\beta^R < u^R < \phi_{0j} - X^R\beta^R \tag{23}$$

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:

$$\int_{-\infty}^{\infty} \int_{\phi_{0,j-1} - X^R\beta^R}^{\phi_{0j} - X^R\beta^R} P(Y = y|u^R, \xi) \prod_{l=1}^{L} P(Y^l = y^l|u^R, \xi) f(u^R|\xi) \mathrm{d}u^R \frac{1}{\sigma_\xi} \phi(\frac{\xi}{\sigma_\xi}) \mathrm{d}\xi \tag{24}$$

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 (24). The conditional probabilities in (24) follow from (21) and (22), 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 $\tau_{\ldots}$ are given by (19) and (20) (and depend on ξ and u^R).

Table E.1. Estimates of Vignette Equations

Vignette	Coef.	s.e.
1	0.000	
2	-1.273	0.065
3	0.705	0.040
4	0.532	0.031
5	1.228	0.064
6	0.297	0.024
7	0.746	0.038
8	1.566	0.077
9	1.527	0.076
10	1.340	0.067
11	-0.945	0.048
12	0.674	0.036
13	1.132	0.058
14	0.479	0.029
15	0.782	0.043
σ_v	0.507	0.023
δ	-0.007	0.019

E Estimates of Vignette Dummies

The dummy coefficients in Table E.1 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. The estimate of δ, the coefficient of the dummy for a female vignette name is small and insignificant. 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).