UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche "Marco Fanno"

DYNAMICS OF WORK DISABILITY REPORTING IN EUROPE

VIOLA ANGELINI
University of Padova

DANILO CAVAPOZZI
University of Padova

OMAR PACCAGNELLA
University of Padova

December 2009

"MARCO FANNO" WORKING PAPER N.104

Dynamics of work disability reporting in Europe*

Viola Angelini[†], Danilo Cavapozzi[‡] and Omar Paccagnella[§]

University of Padua

December 16, 2009

Abstract

In this paper we investigate the role of response styles in the dynamics of work disability reporting. Using the 2004 and 2006 waves of the Survey of Health, Ageing and Retirement in Europe (SHARE), we document that in Europe surprisingly large fractions of individuals change their self-reported disability status within two years. We find that this dynamics can be largely explained by the fact that respondents change the way they evaluate the severity of work disability problems over time.

JEL classification: I10, J14, C33.
Keywords: Work disability, vignettes, reporting heterogeneity.

*We are grateful to Arthur van Soest, Teresa Bago d'Uva, Silvana Robone and participants at the COMPARE Final Policy Meeting (Brussels, 2008), the 7th International Conference on Multilevel Analysis (Amsterdam, 2009), the 7th World Congress on Health Economics (Beijing, 2009) and the 2nd SHARE User Conference (Mainz, 2009) for helpful comments and suggestions. Financial support from the European Commission through the STREP project COMPARE n. 028857 is gratefully acknowledged. This paper uses data from SHARE release 2.2.0, as of August 19th 2009. SHARE data collection in 2004-2007 was primarily funded by the European Commission through its 5th and 6th framework programmes (project numbers QLK6-CT-2001- 00360; RII-CT- 2006-062193; CIT5-CT-2005-028857). Additional funding by the US National Institute on Aging (grant numbers U01 AG09740-13S2; P01 AG005842; P01 AG08291; P30 AG12815; Y1-AG-4553-01; OGHA 04-064; R21 AG025169) as well as by various national sources is gratefully acknowledged (see http://www.share-project.org for a full list of funding institutions).

[†]Department of Economics, Via del Santo 33, 35123 Padova, Italy. E-mail: viola.angelini@unipd.it.

[‡]Department of Economics, Via del Santo 33, 35123 Padova, Italy. E-mail: danilo.cavapozzi@unipd.it.

[§]Department of Statistical Sciences, Via Cesare Battisti 241/243, 35121 Padova, Italy. E-mail: omar.paccagnella@unipd.it.

1 Introduction

The sustainability of European Social Security systems is challenged by the process of population ageing characterizing European societies. Low fertility rates and increasing life-expectancy reduce the proportion of working age individuals in the population and increase the expenditure for the provision of social benefits to the elderly. The design of policies targeted to work-disabled persons plays a prominent role when pursuing the financial equilibrium of modern welfare states. An efficient Social Security system should both maximize the social safety provision needed to offset the income losses due to limited working capacity and, at the same time, prevent potential misuses of disability programs by enrolling only individuals actually unable to participate in the labour market.

While many papers have focused on the determinants of reporting work related health problems, much less work has been devoted to the analysis of its dynamics. One such study (Kapteyn et al., 2008) documents that in the US surprisingly large fractions of individuals change their self-assessed work disability status from one year to the next. Kapteyn et al. (2008) explain this variation in work disability reporting over time with a similar variation in the presence and prevalence of the experience of pain.

In this paper we propose a complementary explanation for the dynamics in work disability reporting, that is the within-person variation in response scales over time. We draw data from the 2004 and 2006 waves of the Survey of Health, Ageing and Retirement in Europe (SHARE), which collect extensive information on health, socioeconomic status and family interactions of individuals aged 50 and over in several European countries, ranging from Scandinavia to the Mediterranean. Our measure of work disability is derived from individual self-evaluations, which have the advantage of summarizing in a single index the variety of factors determining the disability status. Indeed, this condition depends on the overall socioeconomic condition of an individual, the presence of physical and mental problems and their interactions with the labour demand the individual is faced with. However, one problem with using a subjective indicator is that response scales might differ both across individuals and over time. This heterogeneity in reporting styles is called *differential item functioning* (DIF) and brings about the incomparability of self-assessments. With cross-sectional data, several papers (Kapteyn et al., 2007; King et al., 2004; King and Wand, 2007; Paccagnella, 2008) show how anchoring vignettes can be used to control for DIF and assess whether differences in disability rates across countries and socio-economic groups are genuine or they just reflect differences in response scales. Vignette questions ask respondents to evaluate, on the same scale on which they evaluate themselves, the severity of work disability problems of hypotetical scenarios and people and thus help to identify interpersonal and intercultural variations in response styles. When panel data are available, *time invariant* heterogeneity can be taken into account by using conventional fixed or random effect models (Kapteyn et al., 2008).

In this paper we combine the vignette approach with conventional panel data techniques to control for *time variant* heterogeneity and to investigate to what extent individual reporting styles are stable over time. The reason for doing so is that a possible explanation for the within-person variation in work disability reporting is that individual response scales might change over time. The contribution of this paper is twofold. First, we extend the standard hopit model used for estimating vignettes to the longitudinal dimension to allow response scales to vary over time and control for individual unobserved heterogeneity. Second, we apply the model to work-limiting disability and show that reporting styles are not constant over time. We then argue that this within-person heterogeneity in response scales over time explains a large part of the dynamics in work disability reporting observed in the data.

The remainder of the paper is organised as follows. Section 2 briefly reviews the existing literature on work disability. Section 3 introduces the econometrics models. Section 4 presents some descriptive statistics of the data used in this analysis, while in Section 5 we report the results of the empirical analysis. Finally, in Section 6 conclusions and some final comments are reported.

2 Literature review

Self-reported rates of work disability are key measures in many economic and social research areas. As it is well-known, the proportion of elderly individuals in European society is steadily rising and this demographic trend poses the necessity of designing a welfare state able to foster their social and economic inclusion. Since the normal retirement age is gradually increasing throughout Europe, disability rates may constitute an alternative path for early exits from the labour force. In fact, enrolment in disability programs often entails the permanent receipt of benefits and may act as a bridge towards the eligibility to classical labour retirement schemes. Börsch-Supan (2005) shows the wide variability in the disability insurance enrolment rate across European countries. Whereas it exceeds 14 percent in Denmark, Sweden and the Netherlands, it is less than 3 percent in Austria and Greece. These differentials remain unchanged even controlling for cross-country heterogeneity in gender, age and health distribution. A possible explanation for this puzzle may lie in the institutional differences in terms of eligibility rules that make enrolment in disability programs more generous in some countries than in others. Higher allowances and looser health requirements may constitute a financial incentive towards applying for disability insurance in order to opt out of the labour market before the statutory age for old-age pension entitlements. To this end, Burkhauser and Daly (2002) emphasize that receiving disability allowances does not automatically imply the inability of carrying out a job.

Another strand of the literature uses survey data on work disability and anchoring vignettes to correct for differences in reporting styles across countries and socio-economic groups. Kapteyn et al. (2007) show that disability rates based on self-assessments are much higher in the Netherlands than in the US, although health indicators depict the Dutch population as healthier[1]. They use anchoring vignettes to disentangle differences in reporting styles from actual differences in disability rates. Descriptive statistics of vignette evaluations show that US respondents actually have a greater tendency towards using the extreme labelling in the assessments of work-disability. Allowing for such heterogeneity in reporting styles leads to reduce by one half the raw differential in the disability rates between these two countries. In a related work, Van Soest et al. (2007) look at the role of social interactions in the determination of the thresholds used when self-assessing the presence of work-disability conditions. They find that an increase in the fraction of disability benefits recipients in one's reference group does increase the probability of self-rating as disabled. In other words, the disability rate of one's reference group affects the norm underlying an individual's own concept of disability and, as a consequence, the response scale used for her self-assessments. Given the same health condition, individuals living in different social contexts may then express different self-evaluations of their disability status.

While many papers focus on the determinants of reporting work limitations due to health problems, little work has been done so far to analyse how self-assessed work-disability changes over time (Kapteyn et al., 2008) and whether these changes are related to within-person variations in response scales.

[1] A similar research question is addressed in Banks et al. (2007).

3 The longitudinal hopit model

Anchoring vignettes were first introduced by King et al. (2004) for analyzing ordinal survey responses taking into account individual differences in the interpretation of the survey questions. Vignettes are indeed a new tool for enhancing self-report data comparability across countries, communities, social groups, etc. Once individual self-assessments about a particular concept have been collected, respondents are asked to rate a series of situations (i.e. vignettes) describing hypothetical persons dealing with different extents of this concept. Under the assumption that the extent of the situation described in the vignettes is perceived by respondents in the same way (*vignette equivalence*), variability in vignette evaluations is only due to the different reporting styles adopted. Hence, if the same response style is used for both self-ratings and vignette evaluations (*response consistency*), the additional information provided by vignettes acts as an *anchor* to adjust the self-assessments of different individuals according to a homogenous classification allowing for inter-personal comparisons.

In this section we extend the standard hopit model in order to allow the unobserved part of work disability to be correlated over time. We present the general case in which the same individual is observed at T different points in time, although in the empirical application we will analyse only two periods. For each of the time periods we define an equation for the self-evaluation component and an equation for the vignette component. As in the standard hopit model, we assume that the error terms between components are uncorrelated, while the error terms within components are correlated over time.

3.1 The self-assessment component

Let us denote with $t = 1, 2, ..., T$ the time periods and let Y_{it}^* be the own level of the work disability *perceived* by respondent i, $i = 1, 2, ..., n$ at time t. This unobserved response Y_{it}^* is modelled as

$$Y_{it}^* = X_{it}\beta_t + \varepsilon_{it}$$
$$\varepsilon_{it} = \eta_i + \omega_{it}$$

where X_{it} are time-variant exogenous variables and β_t is the vector of parameters to be estimated (for parameter identification no constant is defined). It is important to note that the unobserved term ε_{it} splits in the individual specific component η_i and the idiosyncratic error ω_{it}. We also impose the following identifying assumptions:

$$\eta_i \sim N(0, \sigma^2) \tag{1}$$

$$\omega_{it} \sim N(0, 1) \tag{2}$$

$$\eta_i \perp \omega_{it} \tag{3}$$

$$\omega_{it} \text{ i.i.d., } i = 1, ..., n \text{ and } t = 1, ..., T \tag{4}$$

$$V(\varepsilon_{it}) = \sigma^2 + 1 \tag{5}$$

$$Cov(\varepsilon_{it}, \varepsilon_{is}) = \sigma^2, \quad t \neq s \tag{6}$$

In particular, assumption 3 means that the individual specific effect η_i and the idiosyncratic error ω_{it} are independent and classifies our model as a random effects specification.

Respondent i at time t turns the continuous unobserved Y_{it}^* into a *reported* category Y_{it}, by means of a threshold model with individual-specific thresholds τ_{it}^k

$$Y_{it} = k \quad if \quad \tau_{it}^{k-1} \leq Y_{it}^* \leq \tau_{it}^k, \quad k = 1, \cdots, K$$

where $-\infty = \tau_{it}^0 < \tau_{it}^1 < \ldots < \tau_{it}^K = \infty$. The thresholds are modelled as a function of exogenous variables V_{it} and a vector of parameters γ_t:

$$\begin{aligned}\tau_{it}^1 &= \gamma_t^1 V_{it} \\ \tau_{it}^k &= \tau_{it}^{k-1} + exp\left(\gamma_t^k V_{it}\right) \qquad k = 2, \cdots, K-1\end{aligned}$$

The set of V_{it} may overlap X_{it}, and may also include wave dummies for $t > 1$. The exponential assumption guarantees that the thresholds increase with k.

Following King et al. (1994) and the standard ordered probit theory, the likelihood for the self-assessment component can be written as

$$L_t(\beta_t, \sigma^2, \gamma_t | Y) = \prod_{i=1}^{n} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{k=1}^{K} \left[F\left(\tau_{it}^k | X_i\beta_t + \eta_i, 1\right) - F\left(\tau_{it}^{k-1} | X_i\beta_t + \eta_i, 1\right) \right]^{I(y_{it}=k)}$$
$$\times N(0, \sigma^2) d\eta_i$$

where $F(\cdot)$ is the normal cumulative distribution function and $\gamma_t = (\gamma_t^1, \ldots, \gamma_t^{K-1})$. Since the expression between squared brackets depends on the random effect η_i, we average the likelihood over η_i making use of the normality assumption 1.

3.2 The vignette component

Let j $(j = 1, 2, \ldots, J)$ index vignettes. The unobserved level of work disability described in vignette j as perceived by individual i at time t is modelled as

$$Z_{ijt}^* = \theta_{jt} + \theta_{fem} Female_{ijt} + u_{ijt} \tag{7}$$
$$u_{ijt} = \varsigma_i + \nu_{ijt}$$

where $Female_{ijt}$ is a dummy variable equal to 1 when the person described in vignette j is a female. The gender-vignette variable is introduced by Kapteyn et al. (2007) who argue that respondents can perceive the vignettes differently depending on whether the hypothetical individual is male or female.

Analogously to what we did before, we split the unobserved component u_{ijt} in an individual specific random effect ς_i and an idiosyncratic error ν_{ijt}. We also impose the following identifying assumptions:

$$\varsigma_i \sim N(0, \rho^2) \tag{8}$$
$$\nu_{ijt} \sim N(0, \delta_{jt}^2) \tag{9}$$
$$\varsigma_i \perp \nu_{ijt}, \eta_i, \omega_{it} \tag{10}$$
$$\nu_{ijt} \text{ i.i.d., } \; i = 1, ..., n, \; j = 1, ..., J, \; t = 1, ..., T \tag{11}$$
$$\nu_{ijt} \perp \eta_i, \omega_{it} \tag{12}$$
$$V(u_{ijt}) = \rho^2 + \delta_{jt}^2 \tag{13}$$
$$Cov(u_{ijt}, u_{ijs}) = \rho^2, \quad j = 1, ..., J, \; t = 1, ..., T \tag{14}$$

Respondent i at time t turns the continuous unobserved level of vignette j into a reported category Z_{ijt}, by means of a threshold model with individual-specific thresholds τ_{it}^{k}

$$Z_{ijt} = k \quad if \quad \tau_{it}^{k-1} \leq Z_{ijt}^{*} \leq \tau_{it}^{k}, \quad k = 1, ..., K$$

According to the response consistency assumption, the thresholds τ_{it}^{k} are the same of the self-assessment equation.

It is worth stressing that the equations determining the perceived level of individuals' own work disability and the one of the persons described in the anchoring vignettes contain two distinct random effects. This strategy is aimed at allowing for the fact that individuals may have different perceptions of the work disability limitations when assessing their own status and the vignettes. In other words, the same health limitation may be associated to different work-disability evaluations depending on whether this condition is experienced by the respondent or not.

Preserving the notation used for the self-assessment case, the likelihood for the vignette component can be written as

$$L_t(\theta_t, \rho^2, \delta_{jt}^2, \gamma_t|Z) = \prod_{i=1}^{n} \int_{-\infty}^{+\infty} \prod_{t=1}^{T} \prod_{j=1}^{J} \prod_{k=1}^{K} \Big[F\left(\tau_{it}^{k}|\theta_{jt} + \theta_{fem} Female_{ijt} + \varsigma_i, \delta_{jt}^2\right) - F\left(\tau_{it}^{k-1}|\theta_{jt} + \theta_{fem} Female_{ijt} + \varsigma_i, \delta_{jt}^2\right)\Big]^{I(z_{ijt}=k)} \times N(0, \rho^2) d\varsigma_i$$

where $\theta_t = (\theta_{1t}, \ldots, \theta_{Lt}, \theta_{fem})$. Even in this case, we deal with a conditional likelihood depending on an unobserved random effect. We integrate the random component ς_i out by exploiting the normality assumption 8.

Since self-reported and vignette questions are asked on the same scale, they can be used to jointly identify the β_t, γ_t and θ_t parameters. In particular, the γ_t and θ_t parameters can be identified (up to scale and location normalization) from the vignette equation alone, while β_t parameters can be identified in addition from the self-assessment equation. This also means that in order to identify these parameters, we need that respondents answer to the self-evaluation question as well as to (at least) one vignette question.

Under the independence assumptions 10 and 12, the complete likelihood of the longitudinal hopit model results from the product of the likelihood functions derived for the self-assessment and the vignette component:

$$L_t(\beta_t, \sigma^2, \gamma_t, \theta_t, \rho^2, \delta_t^2 | Y, Z) = L_t(\beta_t, \sigma^2, \gamma_t | Y) \cdot L_t(\theta_t, \rho^2, \delta_{jt}^2, \gamma_t | Z).$$

Parameter estimates are derived by maximizing the likelihood function and integrating out the unobserved effects (η_i and ς_i) via adaptive quadrature techniques.

4 Data and descriptive statistics

Data are drawn from the 2004 and 2006 waves of the Survey of Health, Ageing and Retirement in Europe (SHARE). This survey collects extensive information on health, socioeconomic status and family interactions of individuals aged 50 and over in several European countries, ranging from Scandinavia to the Mediterranean. As part of the COMPARE project, in eight countries (Sweden, Germany, the Netherlands, Belgium, France, Spain, Italy and Greece) respondents aged 50-64 and their spouses are asked to report the presence and severity of problems reducing their working capacity and to evaluate the work disability level of hypothetical persons of whom they get a short description of the health conditions (the so-called "vignettes"). The interesting feature of these data is that the self-reported work disability question and three of the vignettes are asked to the same respondents twice, the first time in 2004 and the second time in 2006.

The self-evaluation question is "*Do you have any impairment or health problem that limits the amount or kind of work you can do?*", while the vignette questions that were asked in both waves are:

1. "[Kevin] suffers from back pain that causes stiffness in his back especially at work but is relieved with low doses of medication. He does not have any pains other than this generalized discomfort."

2. "[Anthony] generally enjoys his work. He gets depressed every 3 weeks for a day or two and loses interest in what he usually enjoys but is able to carry on with his day-to-day activities on the job."

3. "[Eve] has had heart problems in the past and she has been told to watch her cholesterol level. Sometimes if she feels stressed at work she feels pain in her chest and occasionally in her arms."

How much is [Kevin/Anthony/Eve] limited in the kind or amount of work [he/she] could do?

Both for the self-evaluation and the vignettes the possible answers are "*none*", "*mild*", "*moderate*", "*severe*" and "*extreme*".

For identification (see Section 3) we keep only those respondents who have answered both to the self-evaluation question and to at least one vignette in each wave. We also excluded from our analysis Greece and Sweden because at the time of writing of this paper data for these two countries are not yet available. Our final sample is composed by 796 individuals, each of them observed twice. Given that for each respondent we also have at least one vignette evaluation per year, the total number of observations on which the estimates are carried out is equal to 6,356.

As reported in Table 4, the respondents in our sample are prevalently females (56%), living with a partner (84%) and low educated (46%). Since the same individuals are interviewed at two different points in time, the age distribution shifts to the right from one wave to the next: in 2006 respondents are on average two years older. As regards health indicators, most of them tend to be stable over time but the percentages of respondents reporting symptoms of serious health conditions and the presence of chronic diseases increase between 2004 and 2006. This observation is consistent with the fact that health tends to deteriorate with age.

Table 1 and Figures 1 to 4 show the differences between the two waves in the answers provided by respondents to identical self-assessed and vignette questions on work disability. Following the literature, we define a person as worked disabled if her work limitations are moderate, severe or extreme. In general, the rates of self-reported work disability are fairly stable across waves, with more marked differences in Italy and Spain. However, looking at the aggregate rates hides the dynamics: indeed, although in general the differences between waves in the rates of self-reported work disability might not seem large, we still observe substantial transitions in and out of disability (Table 2). The most striking result is that 15.33% of respondents change their self reported work disability status from one wave to the next. This percentage is non-negligible since the time span between the two waves is only two years. In particular, 7.54% of respondents move into disability and the remaining 7.79% move out of it within two years. These percentages are particularly high in Germany and Spain. If we consider the original 5-point scale, the percentage of respondents who change answer categories rises even to 41.83%.

By looking at vignette evaluations at a pure descriptive level, we find evidence of the fact that these variations in work disability reporting might be partially explained by the within-person variation in response scales over time. Table 3 shows that in all countries quite a large fraction of respondents change their evaluation of the work limitations of the hypothetical persons described in the vignettes over time: overall this figure is 42.03% for Kevin, 38.54% for Anthony and 27.65% for Eve. We interpret these results as *prima facie* evidence of the fact that individual response styles cannot be considered as stable over time.

5 Results

In our estimates we consider a large number of factors associated with work disability: demographic characteristics (age and gender), education, health (number of chronic diseases, limitations with mobility, ADL, IADL, obesity, having depression symptoms, a measure of handgrip strength, etc.), marital status and cognitive abilities, as well as country dummies (Germany is the reference country).

Table 5 reports the results of the estimation of the longitudinal hopit model. In both waves the countries with the lowest perceived levels of self-reported work disability are Spain, Italy and France. As expected, the most important determinants of work disability reporting are health conditions, while age and socio-economic variables do not seem to play a role. The results for the threshold equations are presented in columns 2 to 5. Reporting styles significantly vary with the country of residence, health, education and age. Therefore, even controlling for individual random effects, the role played by response scale heterogeneity should not be neglected. A formal Wald test strongly rejects the hypothesis of threshold invariance across individuals[2]. In Table 6 we report the estimates for the parameters in the vignette equations that indicate the level of work disability for the hypothetical persons described in the vignette as perceived by respondents in the two waves (θ_{1t}, θ_{2t} and θ_{3t}) Anthony's vignette is always associated with the lowest level of work disability, while the health conditions of Eve are considered the most work limiting. The estimates also show that the gender of the person described in the vignette matters: the same health conditions are considered less severe for a woman than for a man (θ_{fem} is negative)[3]. This finding may be ascribed to the fact that, *ceteris paribus*, men are more likely to have jobs that are physically demanding and even

[2] The null hypothesis of the test is that all the explanatory variables in the threshold equations but the constant are equal to 0. The Wald statistic is equal to 452.55. The test is asymptotically distributed as a χ^2_{180}, p-value=0.000.

[3] The gender randomization of the hypothetical person described in the vignette is available only in 2004, so we cannot estimate this parameter for 2006.

a mild health problem may hamper their working activity more than in the case of women.

The longitudinal hopit model allows disentangling individual specific random effects from idiosyncratic errors in both the self-assessment and the vignette component. As displayed in Table 6, the presence of individual unobserved factors is likely to play a major role in the determination of both the work disability level and the vignette evaluation. According to our estimates, about 33 percent[4] of the overall unexplained variance can be ascribed to η_i. Hence, neglecting the presence of such random effect would produce an efficiency loss in the estimation. Moreover, individual random effects should not be neglected even in the vignette component, where the fraction of unexplained variance imputable to the random effect ς_i ranges between 22 and 33 percent[5].

We now turn to the main research question of our paper, which is whether response scales can be considered as constant over time. In columns 2 to 5 of Table 7 we report the results of Wald tests for the presence of differences between waves in the parameter estimates of the threshold equations, where the null hypothesis is that the parameters on each explanatory variable are time invariant. Several health conditions, such as having limitations with instrumental activity of daily living, arthritis, health symptoms and chronic diseases have a time-varying effect on response scales. The same holds true for the country of residence, household size and cognitive abilities. In the last row of Table 7 we report the results of joint Wald tests for the null hypothesis of no cross-wave differences in the estimated parameters for all the explanatory variables in the model. For the parameters in the threshold equations, the null hypothesis is rejected in all the four cases. This evidence clearly supports our claim that individual response scales might vary over time. This finding implies that, given the same health status, one person might rate herself as work-disabled in one year but not in the next. In Figure 5 we give a graphical representation of the medians of the individual thresholds in the two waves. The figure shows that the thresholds in 2004 and in 2006 differ quite substantially. In particular, in 2006 the thresholds seem to be shifted to the right: this implies that, as people age, on average they tend to consider the same health problem less limiting for the amount and type of work that they can do. Therefore, some of the transitions out of disability that we observe in the data are not genuine transitions but they just reflect the within-person variation of response

[4] As discussed in Section 3, equation 5 states that the overall unexplained variance of the self-assessment component is $1+\sigma^2$, where σ^2 is the variance of the individual random effect η_i. Therefore the fraction of the overall unexplained variance that can be ascribed to η_i is $\frac{\sigma^2}{1+\sigma^2}$.

[5] According to equation 13, the fraction of unexplained variance imputable to the random effect ς_i is $\frac{\rho^2}{\rho^2+\delta_{jt}^2}$.

styles over time. In some cases this effect might be counterbalanced by the fact that, as people age, their health tends to deteriorate. To quantify this phenomenon, we can carry out counterfactual simulations and estimate how many respondents would move in and out of disability between the two years if they used the same response scales over time. The results are reported in Tables 8 and 9, where for each respondent in our sample we keep constant the response style adopted in 2004 and 2006 respectively. The proportion of respondents reporting a work disability is higher when using the 2004 response scales because, as stressed before, in this case on average individuals have a greater propensity to consider their health problems as work limiting. If individuals kept a consistent reporting style over time, the percentage of respondents moving out of disability would decrease substantially, from 7.74% in the raw data to 2.14% and 0.63% when considering the 2004 and 2006 reporting styles respectively. Therefore, the within-person heterogeneity of response scales over time can largely explain the dynamics in work-disability reporting.

6 Conclusions

In this paper we analyse the dynamics of work disability reporting using data from the 2004 and 2006 waves of the Survey of Health, Ageing and Retirement in Europe (SHARE). We find that a substantial fraction of respondents in our sample change their self-reported work disability status from one wave to the next. In particular, 7.79% of respondents move out of disability, while 7.54% of them become work disabled from 2004 to 2006, according to the self-evaluations. While the transitions into disability can be reconciled with the fact that, as people age, their health deteriorates, it seems more difficult to explain the transitions out of disability, whose percentage is non-negligible if we take into account the fact that the average time span between interviews is only two years. Our evidence is in line with that of Kapteyn et al. (2008), who document that in the US surprisingly large fractions of individuals change their self-assessed work disability status from one year to the next. Kaptyen et al. (2008) explain this variation in work disability reporting over time with a similar variation in the presence and prevalence of the experience of pain.

In our paper we propose a complementary explanation for the dynamics in work disability reporting, that is the within-person variation in response scales over time. In our paper we combine the vignette approach with conventional panel data techniques to control for *time variant* heterogeneity and to investigate to what extent individual reporting styles are stable over time. Recently, many papers have investigated self-reported work disability using vignettes in order to correct for cross-country differences in reporting styles (Banks et al., 2007; Kapteyn et al., 2007; Paccagnella, 2008). To the best of our knowledge, this is the first paper to analyse how self-assessed work disability changes over time using a vignette approach. We do so by extending the standard hopit model implementing the vignette approach (King et al., 2004) to the longitudinal dimension and allowing the unobserved part of the process determining work disability reporting to be correlated over time by means of individual specific random effects. In our estimates we consider a large number of factors associated with work disability: demographic characteristics (age and gender), education, marital status, household size, health (number of chronic diseases, limitations with mobility, ADL, IADL, obesity, having depression symptoms, a measure of handgrip strength, etc.) and cognitive abilities, as well as country dummies.

Our results show that individual response scales are time-varying and such heterogeneity helps explaining the dynamics of work disability reporting observed in the data. In particular, if individuals kept a consistent reporting style over time, the percentage of respondents moving out of disability would decrease substantially. Therefore, the within-person heterogeneity of response scales over time can largely explain the dynamics in work-disability reporting. Moreover, according to our estimates, about one third of the unobserved variability in the process determining work-disability reporting can be explained by individual-specific random effects.

Overall, our findings point to the importance of developing suited econometric techniques and survey instruments that combine vignettes with longitudinal models to control for individual time-variant heterogeneity.

References

[1] Aarts, L., and P. de Jong (1996). European Experiences with Disability Policy. In J. Mashaw, V. Reno, R. Burkhauser and M. Berkowitz (Eds.), Disability, Work and Cash Benefits (Upjohn Institute for employment research, Kalamazoo, MI).

[2] Banks, J., A. Kapteyn, J. Smith, and A. van Soest (2007). Work Disability is a Pain in the *****, especially in England, the Netherlands, and the United States. In D. Cutler and D. Wise (Eds.), Health in Older Ages: The Causes and Consequences of Declining Disability Among the Elderly. Chicago: University of Chicago Press.

[3] Börsch-Supan, A. (2005). Work Disability and Health. In A. Börsch-Supan, A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist and G. Weber (Eds.), Health, Ageing and Retirement in

Europe. First Results from the Survey on Health, Ageing and Retirement in Europe, 253–258. Mannheim: MEA.

[4] Börsch-Supan, A., A. Brugiavini, H. Jürges,J. Mackenbach, J. Siegrist and G. Weber (Eds.) (2005), Health, Ageing and Retirement in Europe. First Results from the Survey on Health, Ageing and Retirement in Europe. Mannheim: MEA.

[5] Börsch-Supan, A., and H. Jürges (Eds.) (2005), Health, Ageing and Retirement in Europe - Methodology. Mannheim: MEA.

[6] Börsch-Supan, A., A. Brugiavini, H. Jürges, A. Kapteyn, J. Mackenbach, J. Siegrist and G. Weber (Eds.) (2008), First Results from the Survey of Health, Ageing and Retirement in Europe (2004-2007). Starting the Longitudinal Dimension. Mannheim: MEA.

[7] Burkhauser, R.V., and M.C. Daly (2002). Policy Watch: U.S. Disability Policy in a Changing Environment. Journal of Economic Perspectives, 16(1), 213–24.

[8] Kapteyn, A., J. Smith, and A. van Soest (2007). Vignettes and Self-Reports of Work Disability in the United States and the Netherlands. American Economic Review 97 (1), 461–473.

[9] Kapteyn, A., J. Smith and A. van Soest (2008). Dynamics of work disability and pain. Journal of Health Economics 27 (2), pp. 496-509.

[10] King, G., C. Murray, J. Salomon, and T. A. (2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. American Political Science Review 98 (1), 191–207.

[11] King, G. and J. Wand (2007). Comparing Incomparable Survey Responses: New Tools for Anchoring Vignettes. Political Analysis 15 (1), 46–66.

[12] Paccagnella, O. (2008). Anchoring vignettes with sample selection. University of Padua, *mimeo.*

[13] Van Soest, A., A. Kapteyn, T. Andreyeva, and J. Smith (2006). Self-Reported Disability and Reference Groups. Wr-409, RAND Labour & Population, Working Paper Series.

A Appendix

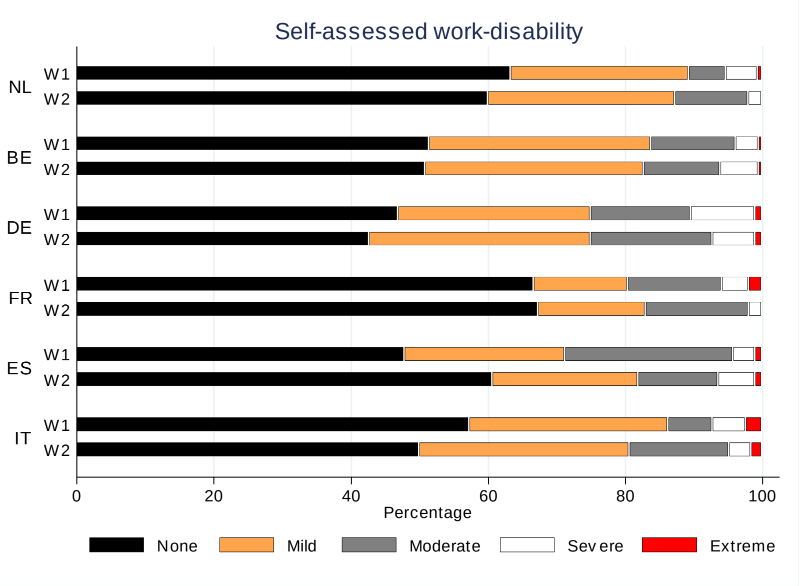

Figure 1: Work-disability self-evaluations by country and wave

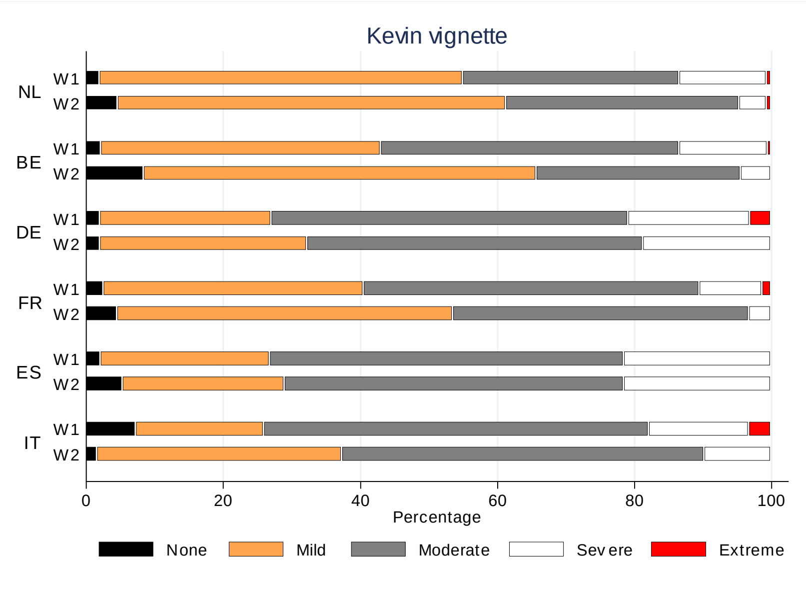

Figure 2: Vignette evaulations by country and wave - Kevin

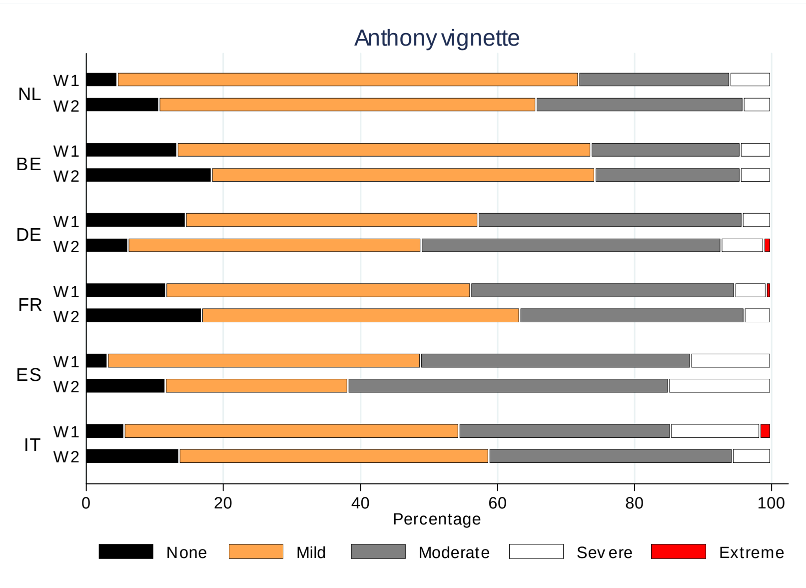

Figure 3: Vignette evaluations by country and wave - Anthony

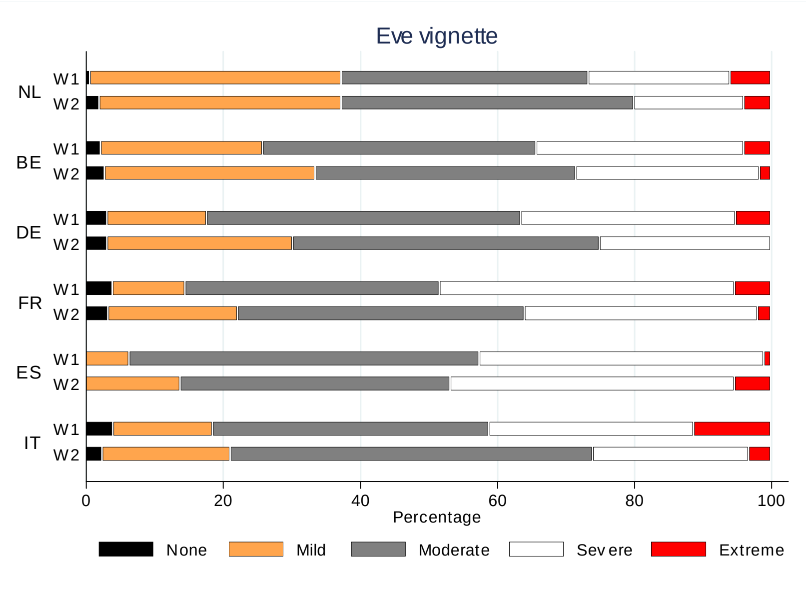

Figure 4: Vignette evaluations by country and wave - Eve

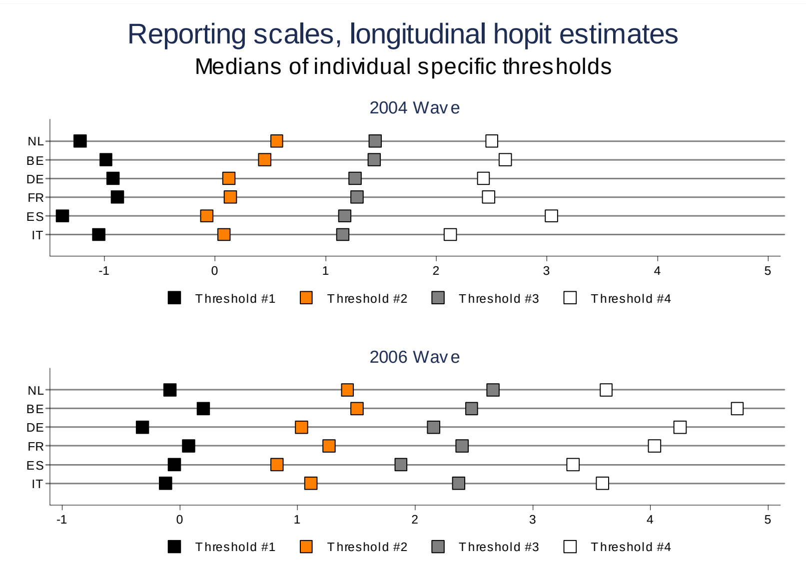

Figure 5: Median thresholds by country and wave

Table 1: Distribution of the answers to the vignette questions in the two waves.

Question	Wave	Answer categories (%) none	mild	moderate	severe	extreme
Self-assessment						
	2004	56.53	25.50	12.06	4.65	1.26
	2006	55.78	26.51	13.19	3.89	0.63
Kevin vignette						
	2004	3.03	35.31	46.28	14.00	1.39
	2006	4.79	45.02	41.36	8.70	0.13
Anthony vignette						
	2004	9.18	52.96	30.57	6.92	0.38
	2006	13.71	47.17	33.08	5.91	0.13
Eve vignette						
	2004	2.40	19.04	40.73	32.28	5.55
	2006	2.39	25.03	42.89	27.04	2.64

Table 2: Transitions in and out disability by country.

Country	Work disability - 2004 wave	Work disability - 2006 wave no	yes
NL	no	84.00	5.33
	yes	3.33	7.33
BE	no	77.09	6.70
	yes	5.59	10.61
DE	no	61.46	13.54
	yes	13.54	11.46
FR	no	75.16	5.23
	yes	7.84	11.76
ES	no	65.96	5.32
	yes	15.96	12.77
IT	no	75.00	11.29
	yes	5.65	8.06
All countries	no	74.50	7.54
	yes	7.79	10.18

Table 3: Vignette evaluations, transitions in and out disability by country.

		2006 wave					
		Kevin		Anthony		Eve	
	2004 wave	no	yes	no	yes	no	yes
NL	no	37.58	17.45	49.66	22.15	20.00	17.33
	yes	24.16	20.81	16.11	12.08	17.33	45.33
BE	no	30.34	12.36	59.22	14.53	15.73	10.11
	yes	35.39	21.91	15.08	11.17	17.98	56.18
DE	no	10.42	16.67	29.17	28.13	9.38	8.33
	yes	21.88	51.04	19.79	22.92	20.83	61.46
FR	no	26.14	14.38	39.22	16.99	5.96	8.61
	yes	27.45	32.03	24.18	19.61	15.89	69.54
ES	no	9.78	17.39	20.21	28.72	1.06	5.32
	yes	19.57	53.26	18.09	32.98	12.77	80.85
IT	no	11.48	14.75	39.02	15.45	5.69	13.01
	yes	26.23	47.54	19.51	26.02	15.45	65.85
All countries	no	23.16	15.19	42.19	19.90	10.61	10.86
	yes	26.84	34.81	18.64	19.27	16.79	61.74

Table 4: Description and sample averages of the variables included in the regressions.

Variable	Description	Sample averages	
		2004 wave	2006 wave
Country of residence			
NL	dummy=1 if the person lives in The Netherlands	0.19	0.19
BE	dummy=1 if the person lives in Belgium	0.22	0.22
DE	dummy=1 if the person lives in Germany (baseline)	0.12	0.12
FR	dummy=1 if the person lives in France	0.19	0.19
ES	dummy=1 if the person lives in Spain	0.12	0.12
IT	dummy=1 if the person lives in Italy	0.16	0.16
Demographics			
male	dummy=1 if the person is male	0.44	0.44
age_54	dummy=1 if the person is aged 54 or less (baseline)	0.40	0.23
age_55_59	dummy=1 if the person is aged between 55 an 59	0.41	0.43
age_60_64	dummy=1 if the person is aged between 60 and 64	0.19	0.34
Education			
low_educ	dummy=1 if the ISCED code is at most 2 (baseline)	0.46	0.46
med_educ	dummy=1 if the ISCED code is 3	0.30	0.30
high_educ	dummy=1 if the ISCED code is at least 4	0.24	0.24
Marital status and household characteristics			
partner	dummy=1 if the person has a cohabiting partner	0.84	0.84
hhsize	household size	2.59	2.39
Physical and mental health			
obese	dummy=1 if the person is obese	0.18	0.19
arthritis	dummy=1 if the person suffers from arthritis	0.19	0.19
chronic	dummy=1 if the person has at least two chronic diseases	0.30	0.32
symptoms	dummy=1 if the person reports at least two symptoms	0.30	0.34
mobility	dummy=1 if the person has at least one limitations with mobility, arm function and fine motor function	0.21	0.20
adl	dummy=1 if the person has at least one limitation with activities of daily living	0.05	0.04
iadl	dummy=1 if the person has at least one limitation with instrumental activities of daily living	0.08	0.07
eurodcat	dummy=1 if the person has depression symptoms according to the EURO-D scale	0.24	0.24
Cognitive abilities and physical tests			
fluency	verbal fluency test score*	2.06	2.08
grip	measure of handgrip strength**	3.72	3.72
Time			
wave2	dummy=1 in the case of 2006 wave	0	1

Note: *In the fluency test respondents are asked to name as many animals as they can think of in one minute. **Two measurements of handgrip strength are taken for each hand and recorded according to a scale spanning from 0 to 100. We take the highest measurement.

Table 5: Longitudinal hopit model, determinants of work disability in the 2004 and 2006 waves.

	Variable	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
2004 wave	NL	-0.599***	-0.282*	0.563***	-0.262**	-0.160
		(0.231)	(0.146)	(0.104)	(0.105)	(0.193)
	BE	-0.293	-0.057	0.322***	-0.120	-0.021
		(0.209)	(0.128)	(0.100)	(0.093)	(0.181)
	FR	-0.775***	0.042	-0.009	0.007	-0.057
		(0.225)	(0.131)	(0.111)	(0.092)	(0.180)
	ES	-0.930***	-0.564***	0.298**	0.070	0.376*
		(0.268)	(0.182)	(0.138)	(0.107)	(0.228)
	IT	-0.975***	-0.192	0.142	-0.063	-0.246
		(0.243)	(0.145)	(0.117)	(0.100)	(0.184)
	male	0.249	-0.114	-0.147*	0.010	0.340**
		(0.183)	(0.110)	(0.081)	(0.080)	(0.156)
	age_55_59	-0.026	0.034	0.024	0.028	-0.033
		(0.135)	(0.084)	(0.062)	(0.060)	(0.123)
	age_60_64	-0.131	-0.124	0.112	0.010	0.054
		(0.174)	(0.111)	(0.080)	(0.077)	(0.156)
	med_educ	-0.209	-0.068	0.090	0.037	-0.152
		(0.153)	(0.094)	(0.067)	(0.067)	(0.130)
	high_educ	-0.231	-0.103	0.124*	0.057	-0.190
		(0.173)	(0.105)	(0.072)	(0.074)	(0.142)
	partner	-0.156	-0.186*	0.039	0.025	-0.030
		(0.183)	(0.110)	(0.080)	(0.083)	(0.162)
	hhsize	-0.014	0.063*	0.003	0.023	0.066
		(0.061)	(0.038)	(0.027)	(0.027)	(0.067)
	obese	0.269*	0.280***	-0.066	-0.255***	0.096
		(0.152)	(0.093)	(0.071)	(0.074)	(0.134)
	arthritis	0.404***	-0.020	0.123*	-0.051	-0.116
		(0.156)	(0.103)	(0.074)	(0.074)	(0.146)
	chronic	0.454***	-0.107	0.066	0.101	0.283**
		(0.141)	(0.090)	(0.063)	(0.065)	(0.138)
	symptoms	0.427***	-0.088	0.084	-0.071	0.260*
		(0.137)	(0.086)	(0.062)	(0.067)	(0.151)
	mobility	0.648***	0.043	-0.102	0.122	0.092
		(0.164)	(0.107)	(0.077)	(0.075)	(0.152)
	adl	0.498*	0.206	0.071	-0.108	-0.284
		(0.258)	(0.166)	(0.120)	(0.134)	(0.244)
	iadl	0.570***	-0.075	0.042	0.060	-0.236
		(0.212)	(0.147)	(0.104)	(0.100)	(0.187)
	eurodcat	0.319**	-0.082	0.024	0.024	-0.172
		(0.144)	(0.092)	(0.068)	(0.069)	(0.151)
	fluency	-0.209**	-0.027	0.008	0.035	0.248***
		(0.102)	(0.058)	(0.042)	(0.045)	(0.089)
	grip	-0.176**	-0.058	0.084**	-0.003	-0.124*
		(0.078)	(0.046)	(0.033)	(0.032)	(0.065)

(See the next page)

	Variable	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
2006 wave	NL	-0.211	0.217	0.201**	0.032	-0.618**
		(0.227)	(0.148)	(0.100)	(0.101)	(0.291)
	BE	0.086	0.497***	0.033	-0.190*	0.200
		(0.209)	(0.138)	(0.095)	(0.099)	(0.311)
	FR	-0.539**	0.362**	-0.088	-0.008	-0.068
		(0.224)	(0.144)	(0.104)	(0.095)	(0.282)
	ES	-0.685***	0.227	-0.364***	-0.091	-0.250
		(0.265)	(0.168)	(0.137)	(0.111)	(0.286)
	IT	-0.516**	0.195	-0.056	0.105	-0.501*
		(0.241)	(0.159)	(0.115)	(0.102)	(0.291)
	male	0.195	-0.040	-0.098	0.169**	-0.363
		(0.173)	(0.098)	(0.070)	(0.083)	(0.237)
	age_55_59	-0.009	-0.188**	0.135**	-0.133*	0.151
		(0.156)	(0.087)	(0.068)	(0.071)	(0.209)
	age_60_64	0.044	-0.224**	0.163**	-0.041	0.049
		(0.171)	(0.097)	(0.074)	(0.076)	(0.216)
	med_educ	0.247*	-0.031	0.163***	-0.056	0.438**
		(0.149)	(0.088)	(0.062)	(0.068)	(0.203)
	high_educ	0.025	0.087	-0.028	0.055	0.101
		(0.167)	(0.092)	(0.070)	(0.073)	(0.201)
	partner	-0.165	-0.113	-0.082	0.034	0.235
		(0.186)	(0.102)	(0.075)	(0.086)	(0.216)
	hhsize	0.101	-0.059	0.057*	0.050	0.067
		(0.072)	(0.044)	(0.033)	(0.032)	(0.108)
	obese	0.155	0.094	-0.004	-0.124*	-0.076
		(0.147)	(0.088)	(0.066)	(0.070)	(0.161)
	arthritis	-0.059	-0.020	0.028	-0.009	-0.124
		(0.161)	(0.098)	(0.073)	(0.077)	(0.191)
	chronic	0.556***	-0.013	0.024	-0.066	-0.064
		(0.142)	(0.084)	(0.064)	(0.067)	(0.176)
	symptoms	0.422***	-0.122	0.121*	0.060	-0.473**
		(0.139)	(0.084)	(0.063)	(0.069)	(0.205)
	mobility	0.682***	-0.076	0.047	-0.033	0.407*
		(0.161)	(0.102)	(0.074)	(0.080)	(0.228)
	adl	0.197	0.018	-0.100	0.002	-0.562*
		(0.294)	(0.182)	(0.146)	(0.147)	(0.301)
	iadl	0.573***	0.246*	-0.240**	0.031	0.162
		(0.219)	(0.131)	(0.110)	(0.105)	(0.255)
	eurodcat	0.656***	0.124	-0.084	-0.014	0.022
		(0.140)	(0.081)	(0.065)	(0.068)	(0.196)
	fluency	-0.149	-0.069	0.055	0.092**	-0.208*
		(0.095)	(0.055)	(0.041)	(0.043)	(0.126)
	grip	-0.147*	-0.061	0.049	-0.075**	0.113
		(0.075)	(0.043)	(0.032)	(0.036)	(0.098)
	wave2		0.986*	0.170	0.164	0.467
			(0.504)	(0.263)	(0.261)	(0.594)
	constant		-0.502	-0.433**	-0.052	-0.132
			(0.373)	(0.186)	(0.178)	(0.339)
Sample size			796			
Log-likelihood			-6611.108			

Note: Standard errors in parentheses, *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table 6: Vignette and variance component estimates.

	2004 wave	2006 wave
Perceived level of work disability		
Kevin vignette (θ_1)	0.539	1.277***
	(0.401)	(0.437)
Anthony vignette (θ_2)	0.070	0.991**
	(0.401)	(0.437)
Eve vignette (θ_3)	1.042***	1.851***
	(0.403)	(0.441)
Female differential (θ_{fem})	-0.077**	
	(0.035)	
Variance of the idiosyncratic error in Kevin's vignette (δ^2_{1t})	0.490	0.385
Variance of the idiosyncratic error in Anthony's vignette (δ^2_{2t})	0.470	0.544
Variance of the idiosyncratic error in Eve's vignette (δ^2_{3t})	0.658	0.640
Variance of the individual-specific effect in the self-assessment component (σ^2)	0.494	
Variance of the individual-specific effect in the vignette component (ρ^2)	0.189	

Note: Standard errors in parentheses, *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table 7: Longitudinal hopit model, testing the cross-wave differences in the parameter estimates.

Variable	Self assessment	γ_1	γ_2	γ_3	γ_4
NL	1.95	7.02***	6.50**	4.17**	1.74
BE	2.26	10.75***	4.50**	0.27	0.38
FR	0.76	3.36*	0.28	0.01	0.00
ES	0.58	12.28***	11.85***	1.10	2.95*
IT	2.47	3.99**	1.50	1.42	0.55
male	0.06	0.30	0.22	1.93	6.18**
age_55_59	0.01	3.75*	1.47	3.05*	0.58
age_60_64	0.62	0.53	0.22	0.22	0.00
med_educ	6.12**	0.11	0.65	0.97	6.01**
high_educ	1.49	2.31	2.36	0.00	1.41
partner	0.00	0.30	1.25	0.01	0.97
hhsize	1.80	5.09**	1.59	0.44	0.00
obese	0.35	2.47	0.41	1.69	0.67
arthritis	4.77**	0.00	0.83	0.16	0.00
chronic	0.29	0.65	0.22	3.25*	2.42
symptoms	0.00	0.08	0.17	1.88	8.31***
mobility	0.02	0.69	1.92	2.02	1.32
adl	0.65	0.61	0.82	0.31	0.51
iadl	0.00	2.76*	3.48*	0.04	1.58
eurodcat	3.02*	3.01*	1.32	0.16	0.61
fluency	0.22	0.31	0.64	0.86	8.74***
grip	0.09	0.00	0.59	2.29	4.06**
constant		3.83*	0.42	0.39	0.62
Total	19.99	40.40**	36.77**	33.31*	37.54**

Note: The Wald tests referring to pairwise comparisons are asymptotically distributed as a χ^2_1. The row *Total* reports the results of joint Wald tests of no cross wave differences for all the parameters in the threshold equations. They are asymptotically distributed as a χ^2_{22} in the self-assessment case and as a χ^2_{23} in the threshold equation case, where the time-intercepts are identified.

Table 8: Longitudinal hopit estimates: transitions in and out disability by country - 2004 thresholds.

Country	Work disability - 2004 wave	Work disability - 2006 wave	
		no	yes
NL	no	88.00	10.67
	yes	0.00	1.33
BE	no	68.72	21.23
	yes	1.68	8.38
DE	no	62.50	19.79
	yes	3.13	14.58
FR	no	80.39	10.46
	yes	3.27	5.88
ES	no	65.96	15.96
	yes	5.32	12.77
IT	no	63.71	28.23
	yes	0.81	7.26
All countries	no	72.74	17.46
	yes	2.14	7.66

Table 9: Longitudinal hopit estimates: transitions in and out disability by country - 2006 thresholds.

Country	Work disability - 2004 wave	Work disability - 2006 wave	
		no	yes
NL	no	97.33	2.67
	yes	0.00	0.00
BE	no	91.06	8.38
	yes	0.56	0.00
DE	no	84.38	11.46
	yes	1.04	3.13
FR	no	94.77	3.92
	yes	0.00	1.31
ES	no	89.36	8.51
	yes	2.13	0.00
IT	no	83.87	15.32
	yes	0.81	0.00
All countries	no	90.83	7.91
	yes	0.63	0.63